Draft Final Report

Peer Review of Michigan State University's PCB Exposure and Effects Studies in the Floodplain of the Kalamazoo River

September 1, 2008

Peer Review Panel

Dr. Ken Dickson, University of North Texas, Peer Review Manager

Dr. Larry Barnthouse, LWB Environmental Services, Inc.

Dr. Anne Fairbrother, Parametrix Inc.

Dr. Keith Grasman, Calvin College

Dr. Mark Harwell, Harwell Gentile & Associates, LLC

Dr. Lyman McDonald, West Inc.

Dr. Charlie Menzie, Exponent Inc.

Dr. William Warren-Hicks, EcoStat, Inc.

Table of Contents

1.0 Introduction

1.1 Background

Following the release of the initial Baseline Ecological Risk Assessment (BERA) (CDM 1999), the Kalamazoo River Study Group (KRSG) provided a series of independent grants to Michigan State University (MSU) for additional ecological studies. The Final Revised Baseline Ecological Risk Assessment (CDM 2003) was finalized before these studies were completed. In February 2007 KRSG voluntarily entered into an Administrative Settlement Agreement and Order on Consent (AOC) for the Site.

The AOC describes a series of supplemental RI/FS activities, potentially including completion of Area-Specific Ecological Risk Assessments. The KRSG requested that the MSU studies be considered as additional lines of evidence for evaluating ecological risks and for subsequent risk management decisions conducted as part of the AOC.

The Statement of Work (SOW) attached to the AOC called for the MSU studies pertaining to floodplain soils to be subjected to a peer-review process prior to incorporation. Floodplain refers to the areas of formerly impounded sediments (i.e., the extent of inundation prior to the lower of water levels in the impoundments). Consequently, the peer review focused on these exposed sediments because USEPA and KRSG have agreed that the aquatic-based ecological food web is unlikely to be the primary risk driver for management of formerly impounded sediments. Dr. John Giesy of Michigan State University (MSU) led the studies on the exposure and effects of PCBs in floodplain soils in the former Trowbridge Impoundment and a reference site (Fort Custer) under independent grants from KRSG. The studies included:

- Productivity assessment of two passerine species and great horned owl;
- Measures of dietary composition for birds;
- Measures of prey tissue PCB concentrations for birds;
- MSU's ecological risk assessment.

The Peer Review Panel (Panel) used the *Final (Revised) Baseline Ecological Risk Assessment Allied Paper, Inc. /Portage Creek/Kalamazoo River Superfund Site* (Site-Wide Baseline ERA; CDM 2003) to provide context and inform the review, but the Panel was not charged with reviewing this document.

However, the Peer Review Panel was charged with independently reviewing the MSU studies and responding to six (6) charge questions, each with supplemental issues noted, and to support answers with citations or other background information as appropriate.

1.2 Peer Review Process

The peer review process was initiated collaboratively by the Kalamazoo River Study Group representing the PRPs and USEPA representing the Trustees by agreeing to engage Dr. Ken Dickson to be the Peer Review Manager. KRSG with the assistance of ARCADIS Inc. and the USEPA prepared independent lists of potential scientists to participate on the Peer Review Panel. Dr. Dickson was requested to constitute the panel but not be constrained by the lists provided. In March 2008 Dr. Dickson recommended to KRSG and USEPA a list of seven (7) scientists to be on the panel. KRSG and USEPA accepted all of the recommended scientists. The Peer Review Panel was constituted in April 2008. The peer review process began on May 13 at a Charge Meeting held in Augusta, Michigan. Participants at the Charge Meeting included representatives from KRSG, ARCADIS Inc., USEPA, MDEQ, MDNR, NOAA, CDM Inc., and MSU. At the meeting Dr. Ken Dickson, the Peer Review Panel, and other participants were introduced to the peer review process to be followed and the project schedule. Dr. Ken Jenkins (ARCADIS Inc.) and Dr. James Chapman (USEPA) provided general background information, introduced the charge to the Panel, and answered questions. The Panel was then briefed by Camp Dresser and McKee (CDM) representatives on the Final Baseline Ecological Risk Assessment (BERA- CDM 2003). Dr. John Giesy, principal investigator of the MSU studies, presented a detailed summary of the MSU studies, results, and conclusions, and responded to questions from the Peer Review Panel and other meeting participants. Representatives from USEPA, MDEQ, and MDNR offered comments/questions about the MSU studies and were requested to provide written comments to the Panel. At the conclusion of the meeting, the Peer Review Panel accepted the charge. The following day, May 14, 2008, the Peer Review Panel and representatives from KRSG, USEPA, and MDEQ visited the study sites used for the MSU studies at the Trowbridge Impoundment and Fort Custer. The group also visited other Kalamazoo River Super Fund Sites (KRSS) along the river.

The Peer Review Panel was provided the following information/documents to inform the review:

- Final (Revised) Baseline Ecological Risk Assessment Allied Paper Inc./Portage Creek/ Kalamazoo River Superfund Site, Michigan Department of Environmental Quality Remediation and Redevelopment Division, April 2003

- Ecological Consequences of PCBs in the Exposed Sediments of Formerly Impounded Areas of the Kalamazoo River – Overview of Studies Conducted by Michigan State University Prepared by Dr. John Giesy and Dr. Matthew Zwiernik – Prepared on behalf of the Kalamazoo River Study Group and the USEPA, May 2008

- The following eight published papers resulting from the Michigan State University studies of the fate and effects of PCBs in the Kalamazoo River:

    - Blankenship, A.L., M.J. Zwiernik, K.K. Coady, D.P. Kay, J.L. Newsted, K. Strause, C. Park, P.W. Bradley, A.M. Neigh, S.D. Millsap, P.D. Jones, and J.P. Giesy. 2005. Differential accumulation of polychlorinated biphenyl congeners in

the terrestrial food web of the Kalamazoo River Superfund site, Michigan. Environmental Science and Technology 39:5954-5963.

- Giesy, J. and M. Zwiernik. 2008. Ecological consequences of PCBs in the exposed sediments of formerly impounded areas of the Kalamazoo River. Presentation to the Kalamazoo River Ecological Risk Studies Peer Review Panel, 13 May 2008.

- Neigh, A.M., M.J. Zwiernik, A.L. Blankenship, P.W. Bradley, D.P. Kay, M.A. MacCarroll, C.S. Park, P.D. Jones, S.D. Millsap, J.W. Newsted, and J.P. Giesy. 2006a. Exposure and multiple lines of evidence assessment of risk for PCBs found in the diets of passerine birds at the Kalamazoo River Superfund site, Michigan. Human and Ecological Risk Assessment 12:924-946.

- Neigh, A.M., M.J. Zwiernik, C.A. Joldersma, A.L. Blankenship, K.D. Strause, S.D. Millsap, J.L. Newsted, and J.P. Giesy. 2007. Reproductive success of passerines exposed to polychlorinated biphenyls through the terrestrial food web of the Kalamazoo River. Ecotoxicology and Environmental Safety 66:107-118.

- Neigh, A.M., M.J. Zwiernik, P.W. Bradley, D.P. Kay, P.D. Jones, R.R. Holem, A.L. Blankenship, K.D. Strause, J.L. Newsted and J.P. Giesy. 2006a. Accumulation of Polychlorinated Biphenyls (PCBs) from Floodplain Soils by Passerine Birds. Environmental Toxicology and Chemistry, Volume 25, pp. 1503-1511.

- Strause, K.D., M.J. Zwiernick, S.H. Im, J.L. Newsted, D.P. Kay, P.W. Bradley, A.L. Blankenship, L.L. Williams, and J.P. Giesy. 2007b. Plasma to egg conversion factor for evaluating polychlorinated biphenyl and DDT exposure in great horned owls and bald eagles. Environmental Toxicology and Chemistry 26:1399-1409.

- Strause, K.D., M.J. Zwiernick, S.H. Im, P.W. Bradley, P.P. Moseley, D.P. Kay, C.S. Park, P.D. Jones, A.L. Blankenship, J.L. Newsted, and J.P. Giesy. 2007a. Risk assessment of great horned owls (*Bubo virginianus*) exposed to polychlorinated biphenyls and DDT along the Kalamazoo River, Michigan, USA. Environmental Toxicology and Chemistry 26:1386-1398.

- Strause, K.D., M.J. Zwiernik , J.L. Newsted, A.M. Neigh, S.D. Millsap, C.S. Park, P.P. Moseley, D.P. Kay, P.W. Bradley, P.D. Jones, A. L. Blankenship, J.G. Sikarskie, and J.P. Giesy. 2008. Risk assessment methodologies for exposure of great horned owls (*Bubo virginianus*) to PCBs on the Kalamazoo River, Michigan. Integrated Environmental Assessment and Management 4:24-40.

- Zwiernik , M.J., K.D. Strause, D.P. Kay, C.S. Park, A.L. Blankenship, and J.P. Giesy. 2007. Site-specific assessments of environmental risk and natural

resource damage based on great horned owls. Human and Ecological Risk Assessment 13:966-985.

- Michigan State University Sampling and Analysis Plan (SAP) and Standard Operating Procedures (SOPs) dated January 7, 2000
- Michigan State University Quality Assurance Project Plan for the Kalamazoo River Area of Concern Baseline Ecological Risk Assessment dated January 7, 2000
- MSU's Kalamazoo Data Base
- NOAA Kalamazoo Data Base
- ARCADIS. 2008. Characteristics of the Formerly Impounded Areas. April 2008.

In addition to the above sources of information the Peer Review Panel requested and received the following:

- MSU's shrew data and a report discussing the results (MSU, 2001);
- MSU's explanation of the process followed to choose the Toxicity Reference Values (TRVs) used in their analyses;
- Information from Michigan Department of Natural Resources and Michigan Department of Environmental Quality on potential future land use management of the formerly inundated floodplains at the sites;
- Comments/Questions from MDEQ and USEPA about the MSU's studies.

1.3 Summary of KRSG and USEPA's Charge to the Panel

1.3.1 General Guidance to Panel Regarding the Charge

The charge to the Peer Review Panel is to review the MSU studies with respect to their suitability as additional lines of evidence for evaluating potential risks to terrestrial receptors exposed to PCBs in floodplain soils in the formerly impounded areas of the Kalamazoo River. A summary of the MSU studies and supporting information was provided to assist the Panel understanding the material to be reviewed. The Panel was also asked to review the Baseline ERA (CDM 2003) for important supporting information and lines of evidence for future risk management decisions. The Baseline ERA (CDM 2003) provided context for the MSU studies, which were designed to provide additional lines of evidence for consideration in the final risk management decisions. However, the Baseline ERA (CDM 2003) was not peer reviewed by the Panel.

The primary objective of the peer review process was for the Panel to provide an independent technical opinion regarding the extent to which the information in the MSU studies could be incorporated as independent lines of evidence, along with those presented in the Baseline ERA (CDM 2003), *in a weight-of-evidence evaluation of ecological risks to terrestrial receptor species* in formerly impounded areas and for subsequent risk management decisions. In reviewing the materials associated with the MSU studies, the Panel was requested to weigh the following general questions when addressing the specific questions presented in the charge:

1) Are the methods employed in the MSU studies appropriate and consistent with the current state of the science and relevant guidance?

2) Have uncertainties associated with the MSU studies been clearly identified and discussed?

3) Do the data and analyses presented in the MSU studies constitute reasonable and appropriate lines of evidence to consider in the evaluation of risks to terrestrial receptors in future risk management decisions?

4) Do the MSU studies represent reasonable and appropriate lines of evidence for consideration in risk management decisions regarding the formerly impounded areas?

1.3.2 Specific Questions to be addressed by the Panel

<u>Exposure Assessments</u>

This section addresses specific issues regarding the evaluation and interpretation of levels of exposure to PCBs for receptors that use the floodplains of the formerly impounded areas. A summary of the types of data and strategies employed by MSU for the evaluation of exposure for the various receptor species is presented in Table 1.1 (included below). The Panel was asked to address the following question regarding exposure and the supplemental issues:

Question 1. What are the relative strengths, limitations, and uncertainties associated with the methods employed by MSU to estimate the exposure of each receptor species to PCBs?

Supplemental Issues to Consider:

1a. Relative strength of various measures of exposure evaluated for each receptor when available individually and in combination. Examples of the types of data MSU considered include: a) literature-,based information on preferred prey; b) Site-specific data on receptor-specific prey items; c) site-specific bioaccumulation factor-based estimates of PCBs in prey; d) direct measures of PCBs in prey; and e) direct measures of PCBs in tissues/eggs of receptors.

1b. Effects of differing dietary preferences on extrapolating from the results of the MSU studies to other species. As an example, how may species-specific dietary preferences of the wrens or bluebirds evaluated in the MSU studies affect extrapolation of risk from these species to robins?

1c. The potential effects of future conditions, such as possible changes in habitat over time due to natural succession or anthropogenic changes to enhance recreational use. Some examples include lowered water table and reduced soil moisture content related to dam removal, transition to meadows including short grass habitat or succession to mature hardwood forest.

Table 1.1 Data Types Available for Refining PCB Exposure Estimates

Available Data	Proposed Use
Bird tissue data presented in Blankenship et al. (2005); Neigh et al. (2006b); Strause et al. (2007a, b, 2008); Zwiernik et al. (2007); and the summary of the MSU studies.	Develop estimate of avian body burden for use in dose model for upper trophic level species.
Shrew and other small mammal tissue data presented in CDM (2003), Blankenship et al. (2005), and the summary of the MSU studies.	Develop estimate of small mammal concentration for use in dose model for upper trophic-level species.
Invertebrate tissue data presented in CDM (2003), Blankenship et al. (2005), and the summary of the MSU studies.	Develop estimate of invertebrate concentration for use in dose model for insectivores.
Egg concentrations from multiple avian species presented in CDM (2003), Neigh et al. (2006b, 2007); Strause et al. (2007a); Zwiernik et al. (2007); and the summary of MSU studies.	Compare to egg-based TRV.
Great horned owl pellet analysis and passerine nestling dietary composition analysis conducted as part of the MSU studies (Neigh et al. 2006a; Strause et al. 2008; Zwiernik et al. 2007).	Refine estimate of dietary composition for purpose of dose modeling.

Effects Assessment

This section addresses specific issues regarding the strategies employed in the MSU studies to evaluate potential effects of PCB exposure on receptors utilizing the floodplains of the formerly impounded areas. The Panel was charged to address the following questions regarding effects and the supplemental issues:

Question 2: What are the relative strengths, limitations, and uncertainties associated with the productivity assessments conducted by MSU on passerines and great horned owls (Neigh et al. 2006a, 2007; Strause et al. 2007a, 2008)?

Supplemental Issues to Consider:

2a. Strengths and limitations of directly measuring productivity in the field compared to extrapolating from controlled laboratory studies.

2b. Extrapolation of results from field productivity studies to other species such as the American robin, which was the receptor species considered in the Baseline ERA (CDM 2003).

2c. Evaluation of potential causal factors (e.g., PCB concentrations, habitat differences, etc.) associated with any difference in measures of productivity in passerines relative to the reference site.

Question 3: What are the relative strengths, limitations, and uncertainties associated with the hazard quotient calculations performed by MSU to evaluate potential risk to passerines, great horned owls, and shrews (Neigh et al. 2007; Strause et al. 2007a, 2008)?

Supplemental Issues to Consider:

3a. Choice of toxicity reference value (TRV), including relevance to receptor species and quality of study (e.g., duration, inclusion of sensitive life stages, exposure range, endpoints measured).

3b. Uncertainty resulting from extrapolating from laboratory study to field.

3c. Uncertainties in extrapolating from one species to another.

Applicability of the Investigations

This section of the charge addresses the overall quality of the data and the analyses presented in the MSU studies and their applicability for the evaluation of ecological risk and supporting risk management decisions for the floodplains of the formerly impounded areas. With this in mind the Panel was asked to address the following questions:

Question 4: What are the relative strengths, limitations, and associated uncertainties that should be considered when evaluating the results of these studies as potential lines of evidence in future risk management decisions?

Supplemental Issues to Consider:

4a. Study designs including (but not limited to) sample size, replication, temporal duration, and aggregation of data.

4b. Data interpretation, including the choice and application of statistical methods.

4c. Approach for addressing natural variability.

4d. Identification and characterization of uncertainties.

4e. Adequacy of the data to support inferences on population-level effects.

Question 5: What are the relative strengths, limitations, and associated uncertainties that should be considered when extrapolating from the results of MSU studies conducted in the former Trowbridge Impoundment to the other formerly impounded areas of the Kalamazoo River?

Supplemental Issues to Consider:

5a. Numeric and spatial distributions of PCBs in floodplains of former impoundments.

5b. Habitat characteristics in floodplains of formerly impounded areas.

5c. Likely utilization of floodplains in formerly impounded area by the receptor species evaluated in MSU studies

Risk Management

This section of the charge addresses the potential usefulness of the MSU studies in supporting risk management decisions for the floodplains in the formerly impounded areas. It is possible that the results of the MSU studies would be incorporated as independent lines of evidence, along with data from the Baseline ERA (CDM 2003), in an Area-specific ecological risk assessment process. With this in mind please address the following question.

Question 6: Please comment on the applicability of the information presented in the MSU studies for informing risk management decisions.

2.0 Panel's Draft Responses to Charge Questions

2.1 General Charge Questions

2.1.1 Are the methods employed in the MSU studies appropriate and consistent with the current state of the science and relevant guidance?

Panel's Draft Response: The MSU studies encompass significant field studies measuring exposure and effects as well as risk assessment models for estimating both, reflecting the current state-of-the-science. This combined strategy has several advantages, including the potential to validate modeled predictions with real-world data for the several species that were actually studied, and the ability to use PCB concentrations in soil and food items to model both exposure and effects for species that were not studied in the field. The general and execution of the field studies were good and comparable in quality to other field assessments of PCBs in riparian habitats in large river systems. However, the methods used by MSU for assessing passerine productivity and for selecting TRVs were not the current state of the science. Another limitation was use of non-rigorous statistical design and analysis of results.

2.1.2 Have uncertainties associated with the MSU studies been clearly identified and discussed?

Panel's Draft Response: The MSU Summary Report and publications resulting from the study do not adequately explain the uncertainties associated with exposure, effects, ecological risks and risk management conclusions. It is a **strength** of the MSU field studies that they provided site-specific data, thereby reducing uncertainties inherent in the approach of using primarily literature-based values to estimate exposures and effects. However, the small sample sizes, issues related to the adequacy of reference sites, and absence of information on some important pathways of exposures **limit** the impact of the field-based approach on reducing uncertainties.
An approach to better explain (and quantify) uncertainties would be to examine the exposure model(s) in the MSU study and BERA using formal uncertainty analysis to explore the plausible range of risks in the system. For example, one of the variables in such a model is the particular diet of an endpoint species. A probability distribution (consistent with field measurements) of dietary sources could be generated allowing quantification of the uncertainty in exposure as a function of diet. . Similarly, use of different specific bioaccumulation factors within the range of plausible values for each could be explored in a set of Monte Carlo simulations using hazard quotients. In general, rather than focus on a single quotient , the risk assessments would be strengthened by presentation of a distribution of risk levels tied to the uncertainties in the underlying data or model structure (e.g., relative importance of different dietary pathways).

2.1.3 Do the data and analyses presented in the MSU studies constitute reasonable and appropriate lines of evidence to consider in the evaluation of Site-specific risks to terrestrial receptors in Area-specific risk assessments?

Panel's Draft Response: It is the Panel's opinion that data developed by the MSU studies have the potential to inform the site specific risks (that is, information about the Trowbridge impoundment) to terrestrial receptors, but the analyses of data were inadequate and in some cases inappropriate to inform area-specific risk assessments (e.g., extrapolation to the other impoundments or the entire site). The MSU data analyses do not take advantage of the information content contained in the data sets, and therefore do not fully explore the lines of evidence that are inherent in the collected information. Thus, while the MSU studies do present information that can reasonably contribute to a multiple-lines-of-evidence assessment, these data, results, and conclusions need to be considered with caution and appropriate recognition of their uncertainties and limitations.

2.1.4 Do the MSU studies represent reasonable and appropriate lines of evidence for consideration in risk management decisions regarding the former impounded areas?

Panel's Draft Response: The results of the MSU studies should be included in a multiple-line-of- evidence approach for risk assessment and risk management decision making, with caution and appropriate recognition of their uncertainties and limitations. The MSU data and the BERA data should be used to develop an integrated multiple-lines-of-evidence-based ecological risk assessment (i.e., possibly using data from both studies in a single data analysis approach) to inform risk management decisions for the formerly impounded area. However, the uncertainties associated with the multiple lines of evidence from the BERA and MSU studies should be identified and formally quantified so that they can be more effectively considered and weighed in the risk management process.

2.2 Specific Charge Questions Addressed by the Panel

2.2.1 Question 1: What are the relative strengths, limitations, and uncertainties associated with the methods employed by MSU to estimate the Site-specific exposure of each receptor species to PCBs?

The **strengths** of the MSU studies lie in their direct measurement of PCB concentrations in prey items, soil, receptor species and in actual diets of receptors of choice for the field studies. These measurements significantly reduce the uncertainty of exposure estimates for the receptors of concern. The **limitations** of the MSU studies include small sample sizes for some species/trophic levels (such as n=6 for earthworms), lack of spatially explicit data (e.g., BAF functions cannot be determined, only BAF constants), and measurement of PCB concentrations as wet weight (which adds significant variability as opposed to reporting out as dry weight). These limitations result in continued uncertainties in dietary estimates of food chain-modeled species, although the uncertainties were significantly reduced as compared to those in the BERA where literature-based BAFs and diets were used to estimate dose. However, a number of other **limitations** lead to **uncertainties** in MSU's risk conclusions, as well as in extrapolations across habitat types, impoundments, and time periods. MSU's papers are limited in their reporting of locations and habitat types associated with samples. Low sample sizes for some variable types reduce the potential for reanalysis of the data on a spatial basis. In other cases, the compositing of data across sampling areas and habitat types has reduced the level of resolution of the exposure analysis. The conceptual model was limited to assessing current risks to a specific set of receptors, and failed to adequately address the pathway that was identified as critical in the BERA. The analysis of PCB weathering is flawed in the calculation of relative potency factors for irrelevant exposure pathways (e.g., MSU's data show that shrews and other small mammals constitute a small fraction of the owls' diet, hence RPs meant to represent weathering of congener mixtures along this pathway are meaningless). Other limitations and uncertainties are identified in the Panel's responses to specific questions herein.

Supplemental Issues to Consider in Question 1:

1a. Relative strength of various measures of exposure evaluated for each receptor when available individually and in combination. Examples of the types of data MSU considered include: a) literature-based information on preferred prey; b) Site-specific data on receptor-specific prey items; c) site-specific bioaccumulation factor-based estimates of PCBs in prey; d) direct measures of PCBs in prey; and e) direct measures of PCBs in tissues/eggs of receptors.

Panel's Draft Response: Site-specific diet data generally are preferable to literature-derived diets. The results from the MSU studies could be used to calculate species-specific diets used in the area-specific ERAs. However, differences in diets between impoundments, as influenced by habitat characteristics, site history, and other factors must be taken into account to the extent possible. Many birds, in particular robins, spend a substantial amount of their foraging time off-site. Exposure estimates for these species should be adjusted to account for off-site foraging. Similarly, some species such as shrews have very small home ranges, so exposure estimates should not be averaged across an entire impoundment. MSU's data also show that a substantial fraction of the diets of bluebirds nesting in the former impoundments consist of aquatic insects, and the great horned owl diets also have a significant amount of items from the aquatic food web (Strausse et al., 2008) Therefore, the potential contribution of PCB exposure from aquatic insects should be considered in area-specific risk assessments of these (or similar) species. Supporting references are: Neigh et al. (2006) and Strausse et al. (2008).

Panel's Draft Response to Question 1a – Modeled (BAF-based) exposures used in BERA vs. site-specific exposures from MSU studies: Site-specific data on bioaccumulation are nearly always preferable to literature-based bioaccumulation factors; however, site-specific BAFs are not necessarily transferable even between nearby sites. Bioavailability of PCBs in soils may differ, and the utilization of the site by target receptors may also vary between sites. When extrapolating to other impoundments, it will be necessary to account (if possible) for uncertainties related to possible variations in PCB bioavailability and receptor dietary preferences. Because the biota-soil accumulation factors (BSAFs) are carbon-normalized, extrapolation among sites is more reliable (assuming there are data on soil carbon for all areas). Similarly, differences in PCB congeners at the various sites should be taken into account when doing such extrapolations. Supporting Reference: Blankenship et al. (2005) contains site-specific biological accumulation factors (BAFs) and biological magnification factors (BMFs) calculated from the Trowbridge impoundment data for birds; by contrast, the BAF for robins used in the BERA was based on a study of herring gulls feeding on alewives.

The Panel is concerned that the method used to determine the BAF factors may overestimate the BAF, resulting in higher "safe" values in the soil when calculating PRGs for clean up. Note that PCB concentrations in soil are expressed on a dry-weight basis, whereas tissue concentrations are all on a wet-weight basis (or lipid normalized, but still wet weight). Since we do not know the percent moisture in any of the biota or tissue samples, we do not know what the actual ***mass*** of PCBs is that has moved up the food chain; the relative concentrations are confounded by the differences in percent moisture among individuals and between species. In addition, the method for preparing biota for PCB analysis may change their hydration, thus altering their measured wet weight. For example, earthworms are depurated prior to chemical analysis by placing them on filter paper for several hours. During this time, they likely desiccate to some extent; thus, their measured "wet weight" is not a true field weight and more variability is added to the analysis (plus individual earthworms will desiccate to different degrees). Although this method of calculating BAFs on a wet-weight basis has become the standard approach for terrestrial systems (but not for aquatic food-chain analyses), it continues to be a major flaw in the way terrestrial food-chain risks and safe soil values are estimated.

Panel's Draft Response – Use of MSU Site-Specific Tissue Data in BERA Dietary Exposure Models: The **strength** of this approach is to provide site-specific BSAFs and BMFs and measured concentrations in biota, rather than basing the food web model primarily on literature-based estimates. This can incorporate soil-specific effects (e.g., soil carbon), congener-specific differences in accumulation rates, and species-specific information related to the site (particularly for raptors, where literature-based data are very sparse). A **limitation** is that the robin diet was not specifically modeled, so it may be that all appropriate biota were not sampled. However there are some data for earthworms, so some estimate of site-specific risk could be developed using the dietary model in the BERA. Also, the recently provided shrew data could be used for site-specific exposure estimations. Background information in the BERA stated that robins can consume up to 90% of their diet as invertebrates during the breeding season, and only up to 20% invertebrates in the remainder of the year. The BERA assumed 51% of the robin's diet is from soil invertebrates. The USEPA Wildlife Exposure Factors Handbook (USEPA 1993) identifies 59 – 71% invertebrates in the robins spring/summer diet. No explanation is provided for why the BERA selected 51% as the amount of soil invertebrates ingested. Robins actually eat a wide variety of soil invertebrates, but as only the earthworm was measured for PCB concentrations, it was considered a "worse case" for PCB uptake and could be used in a site-adjusted estimate of dietary exposure to robins.

Panel's Draft Response – Use the "Bolus" Data from the Avian Nesting Study to Further Verify Dietary Exposure Estimates: The **strength** of this approach is that the food bolus represents precisely what the nestlings are eating. By comparing the concentrations in this bolus to the estimated concentrations from the dietary exposure model, the model can be further refined to accurately reflect the diets and exposures of the studied species. This may provide some additional realism for extrapolating to the non-measured species, such as the robin. The **limitation** of this approach is that only the house wren is truly feeding on only terrestrial foods, while the eastern bluebird, the tree swallow, and the great horned owl access some (or most) of their diets from the aquatic food chain. Thus, relating diet to soil contamination alone is difficult, resulting in substantial uncertainty.

Panel's Draft Response – Great Horned Owl (GHO): The prey item sampling (pellets and prey remains) was a critical part of the ecological studies that provided dietary composition data for bottom-up modeling of PCB exposure. Such studies must pay particular attention to prey identification and quantification in order to avoid biases that over- or under-estimate the frequency of particular food items. There are inconsistencies and incomplete explanations in MSU's descriptions of GHO prey item sampling and analysis. The SOP (273) and two published studies (Strause et al. 2007b, Zwiernik et al. 2007) vary somewhat in the citations for the prey analysis methods. Strause et al. (2007b) appears to have the best description of MSU's actual methods, although more details would be helpful. The methods actually employed differ significantly from those in the SOP, especially with respect to the schedule for collection (4X/month in the SOP vs. 2X per breeding season in the actual study) and level of data aggregation (individual pellets in SOP vs. composite samples in the study). In some cases the cited papers do not appear to provide strong support for the particular point that is being made by MSU (e.g., Hayward et al. 1993 in Zwiernik et al. 2007). While overall the estimated dietary composition methods appears to have provided satisfactory data, the inconsistent and incomplete documentation of methods potentially limits quantitative comparisons to other studies (which may have used other methods) and restricts the ability of other researches to replicate these protocols. The **strengths** of this study are the direct measurements of PCBs in prey items that rarely are analyzed, and a reasonable comparison between KRSS and the reference site. Further **strengths** are the presentation of data on both a mass-basis and a concentration-basis, plus inclusion of both means and 95% UCLs of the means (although the means are geometric means, rather than the more appropriate means based on a log normal distribution).

Panel's Draft Response to MSU Data Types Available for Refining PCB Exposure Estimates (Table 1.1 in Charge):

1. *Use of bird tissue data to develop estimates of avian body burden for use in dose model for upper trophic level species*

Panel's Draft Response: Using the bird tissue data from the MSU studies would be preferable to using literature-derived bioaccumulation factors, as were done in the BERA, but only if a sufficient number of samples exist for the particular species.

2. *Use of shrew and other small mammal tissue data to develop estimate of small mammal concentration for use in dose model for upper trophic level species*

Panel's Draft Response: Use of shrew and other small mammal tissue could potentially provide a more realistic exposure estimate for terrestrial predators, if the sample sizes (number of animals and range of prey species) are sufficient to provide meaningful estimates of prey tissue concentrations. This is comparable to the work done by Strausse et al. (2008) for estimating great horned owl diets.

Panel's Draft Response – MSU Shrew Studies: The MSU shrew study provides the following data:

- PCB tissue levels in shrews from the former Trowbridge impoundment; collections are from four grids and yielded 17 animals for tissue analysis.
- Trapping data from Trowbridge (4 grids) and the reference location at Fort Custer (2 grids); these data provided abundance caught relative to trap nights.

The Panel felt that these data could be used to evaluate exposures and risks to higher trophic level predators as well as risks to shrews, that is, the shrews should be considered both a pathway and receptor species of concern. Each of these is described below.

Panel's Draft Response – Evaluating Exposures to Higher Trophic Levels: Shrews comprise part of the diet of various raptors and mammals. Therefore, the PCB body burden data for shrews can be used to evaluate exposures to these trophic levels. Strengths of the MSU include:

- An adequate sample size of shrew tissues to estimate exposures to higher trophic levels. The panel recommends that this be carried out using a statistical/probabilistic approach. Appropriate statistics should include estimates of sufficient statistics (and associated estimation uncertainty) based on the lognormal distribution.
- Opportunity to develop shrew:soil BAFs from the grids using the tissue data and the average for soils represented by the grid. These shrew:soils BAFs could then be applied to other former impoundments. For soil, the statistic would represent the most likely value at the spatial scale of the sampling grid.

Panel's Draft Response – Limitations/Qualifications of the MSU shrew studies to estimate exposures to higher trophic levels include:

- With respect to future use of the body-burden data, it will be important to examine the spatial relationships between the shrew tissue data and the relevant soil data at the scale of sampling grids. BAF values are not necessarily linear with soil concentrations. Therefore, the change in BAF(s) with soil concentrations can be important and should be examined prior to applying BAFs from one locations to another.

- The Panel could not easily tell which shrews come from which grids and how this relates to the associated soil concentrations. An examination of these geographic and concentration relationships will be important with respect to making use of these data for other parts of Trowbridge (where shrews were not collected) and for other former impoundments.

Panel's Draft Response – Evaluating Exposures to Shrews as Receptors: Measurements of PCBs in soils, earthworms, and shrews could be used to estimate PCB exposure and risks to shrews. Strengths of the MSU studies include:

- Direct measures of PCBs are available for a field-collected food item –earthworms – that likely reflect higher concentrations of PCBs than other potential invertebrate food items for shrews.

- If there is sufficient variability in soil concentrations among the soil-worm collection sites to develop a BAF function beyond a simple constant, then the field-derived data could be applied to other areas. Otherwise, using the highest BAF will provide a conservative estimate.

- The combination of dietary exposure estimates and shrew body burdens provides a basis for a weight-of-evidence assessment of risk to shrews. However, the best use of the shrew data is as input to the food chain model for higher order predators.

Panel's Draft Response – Limitations/Qualifications MSU shrew studies include:

- The Panel could not easily discern the locations from which the earthworms came and the specific calculations used to associate earthworm tissue levels to soil concentrations. In order for the earthworm data to be of value, there needs to be a clear association of tissues with co-located soils. An examination of these geographic and concentration relationships is critically important to making use of these data for other parts of Trowbridge and for other former impoundments.

- The sample size for earthworms – six – is small especially in light heterogeneity in soil PCBs and, therefore, care must be taken on how to extrapolate these data to other regions within Trowbridge and to other former impoundments. In particular, because BAFs may vary with soil concentration, this relationship should be examined as part of a data usability assessment. If the range of soil concentrations used to derive earthworm BAFs for Trowbridge overlaps (is representative) of the range in other former impoundments, there is higher confidence in using the Trowbridge data. If the data sets are very different, there is less confidence. However, if other former impoundments have lower concentrations, the data for Trowbridge could still be used as a bounding analysis. Given the low numbers of earthworm samples, the Panel concurs with MSU to use the higher-derived BAF for extrapolations.

- The Panel recommends that these data be used as part of a weight-of-evidence approach and that the approach not rely on the HQ values that have been calculated. Instead, it is recommended that these types of calculations be repeated after the data have been processed as described above.

- The geometric mean underestimates the "most likely" value of log normally distributed variables (see Figure 2.1). Therefore, exposure estimates should properly reflect an underlying log normal distribution. The most likely value can be approximated by the median of a set of measurements. Other methods for approximating the most likely value of a log normal distribution are available in standard text books.

- The MSU shrew assessment relies on averages and upper bounds on averages for individual impoundments. The impoundments are large relative to foraging areas of shrews. Therefore, a spatially-explicit approach should be used that considers the exposures distributions to individuals shrews across the impoundment(s). Exposure and risks can then be presented in terms of sub-areas or as fractions of the local population. The Spatially-Explicit Exposure Model (SEEM) offers one way for doing this.

3. *Use of invertebrate tissue data to develop estimate of invertebrate concentration for use in dose model for insectivores (birds? shrew?)*

Panel's Draft Response: The available invertebrate tissue data could, in principle, provide more realistic exposure models for insectivorous birds and mammals, if the sample sizes are sufficient to provide meaningful estimates of prey tissue concentrations. It would be most appropriate to use invertebrate data in a spatially explicit analysis of the impoundments, rather than an average concentration for the whole impoundment. Concentrations in these animals may differ significantly with changes in PCB soil concentrations and soil types, both of which may have considerable small-scale spatial heterogeneity. If the data are extrapolated to other impoundments, this potential for spatial differences should be identified as an uncertainty.

4. *Comparison of egg concentrations for multiple avian species to egg-based TRVs.*

Panel's Draft Response: Although the MSU studies have substantially increased the number of measured egg concentrations available for comparison to egg-based TRVs, the TRVs themselves are all literature-derived and of uncertain applicability to species present at the site. Hence, increasing the size of the egg concentration data set may not significantly reduce the overall uncertainty in the egg-based HQs.

5. *Use of great horned owl pellet analysis and passerine nestling dietary composition analysis to refine estimate of dietary composition for purpose of dose modeling.*

Panel's Draft Response: The owl pellet analyses and passerine nestling dietary composition data collected by MSU could provide more realistic dose models; however, it will still be necessary to account for potential variations in diet (1) between impoundments, and (2) between present and potential future conditions. In addition, it must be demonstrated that sample sizes from the MSU studies are sufficient to support their use in dose modeling.

> ***1b. Effects of differing dietary preferences on extrapolating from the results of the MSU studies to other species. As an example, how may species-specific dietary preferences of the wrens or bluebirds evaluated in the MSU studies affect extrapolation of risk from these species to robins?***

Panel's Draft Response -- bluebird/house wren exposure (dietary dose) to other species (especially robin) exposure: Wren and bluebird diets can not be directly extrapolated to other species; however, PCB concentrations in invertebrates and earthworms can be used to estimate dietary doses to other species, using information on dietary preferences of those species.

The Panel suggests that bluebird data may provide a bounding condition. This would help set the limits on the use of these data. Examination and contrast of feeding habits and – if available – information on sensitivity for other birds of this body size and metabolism to PCBs might lead to a conclusion that you can not extrapolate directly but may be able to make inferences. Supporting Reference: Blankenship et al. (2005) reports PCB concentration data for a variety of food items present in the Trowbridge impoundment.

> ***1c. Potential effects of future conditions on risk, such as possible changes in habitat over time due to natural succession or anthropogenic changes to enhance recreational use. Some examples include lowered water table and reduced soil moisture content related to dam removal, transition to meadows including short grass habitat or succession to mature hardwood forest.***

Panel's Draft Response – Time-related extrapolation issues: Future land use, natural succession, and reduced inundation frequency can all be expected to change the utilization of the formerly inundated areas by key receptor species, and possibly also change the diets of those species that utilize these areas. For this reason, conclusions concerning risks to bluebirds, house wrens, robins, and great horned owls reached in the papers published by the MSU group may not hold in the future. However, provided that the influences of future land use, succession, and inundation on habitat suitability and prey availability can be predicted, it may still be possible to use the MSU data to predict foraging, diet composition, and PCB exposures under future conditions. It also is instructive to note that the bluebird study identified disturbed habitat in the Trowbridge impoundment (as compared to the Fort Custer reference area) as one potential reason for reduced productivity. It may be that as succession-related changes occur, or as changes in the hydrologic regime follow from removal of remnant dam structures, habitat is improved, thus resulting in better productivity for these and other passerines.

To facilitate consideration of the effects of natural succession, existing information on succession patterns in the riparian habitat that borders river systems in Michigan should be used to define two to four general succession conditions and associated biota. Habitat characteristics associated with each of these stages could then be used to predict the diets of passerine birds and great horned owls present in each impoundments as functions of succession stage.

Panel's Draft Response – Implications for the Earthworm Pathway: The BERA indicated that the relatively high concentrations of PCBs in earthworms make that the critical pathway for exposures in the Kalamazoo River Basin, showing up in the exposure pathway through the robin. However, the MSU studies did not emphasize this critical pathway and may not have sufficient data to reflect the potential risk to wildlife through earthworms. MSU did not collect data on the robin, indicating that resulted in part from there being few earthworms in the former impoundment areas because of relatively frequent inundation by river water. As noted elsewhere, this condition may not remain the case in future scenarios. The Panel believes that the sparseness of earthworm data (Nn=6) is a major gap in the MSU analysis. That means that the BERA model for the robin has neither been substantiated nor refuted, thereby remaining an important decision analysis point. Even if robins are not currently foraging heavily on-site (because of availability of better foraging habitat in adjacent uplands), small mammals such as shrews are present on site and would be expected to feed on earthworms. Although the former impoundment areas of concern do not seem to be good robin habitat at present, consideration of the soil—earthworm exposure pathway is essential for evaluating risks to other earthworm-eating wildlife such as woodcock, snipe, shrews and robins.

For example, in a risk assessment of a variety of soil contaminants at Oak Ridge National Laboratories (ORNL, Efroymson et al. 1997), soil cleanup goals for PCBs were driven by shrews and then woodcocks, leading to values of 0.371 and 0.655 $mg \cdot kg^{-1}$ soils, respectively, based on a hazard quotient (HQ) model similar to the models used in the Kalamazoo River BERA. Note that the ORNL PCB cleanup goals are significantly below current soil PCB concentrations in the former impoundments on the Kalamazoo River. On the other hand, preliminary cleanup goals for floodplain soil at the Housatonic River site are far higher (21.1 $mg \cdot kg^{-1}$ to 43.5 $mg \cdot kg^{-1}$) based on results of a site-specific study of short-tailed shrew population dynamics (USEPA 2004, GE 2005).

Given the wide range of cleanup values derived for other sites with PCB-contaminated soil, a focus on shrews would seem to be particularly appropriate to address the potential earthworm pathway exposures. MSU's October 05, 2001 update report (MSU, 2001) provides data on tissue PCB concentrations in shrews collected from the former Trowbridge impoundment. Table 4-5 of that report summarizes tissue concentration data for 17 shrews. These data could be used to estimate doses to predators such as foxes and great horned owls that would be expected to prey on shrews. They could also be used to directly address risks to shrews; however, the need to rely on TRVs derived from rat studies would make any such assessments uncertain.

Two plausible future scenarios exist that would likely change the frequency of flooding of these areas, thereby providing much more suitable habitat for earthworms, and consequently increasing the potential PCB exposures to earthworm-eating birds and mammals:

1) The representative from the Michigan Department of Environmental Quality (MDEQ) stated that there remains a firm commitment by the State to complete the removal of the

remnant dam structures, thereby allowing a free-flowing Kalamazoo River. It is apparent that by removing these structures, the River water levels would be lowered by a few feet, particularly in the lower half of the former Trowbridge impoundment upon removal of the remnants of the Trowbridge Dam. In that circumstance, the frequency of inundation of the former impoundment sediments would decrease significantly, likely to the degree that an earthworm population would become established.

2) There is little doubt in the scientific community that global climate change is underway and will continue for the next several decades, as documented most recently in the Intergovernmental Panel on Climate Change (2007) reports. While precipitation in the Great Lakes region may increase, temperatures and evapotranspiration rates are also expected to increase, with a net lowering of water levels being plausible, although uncertainty exists, especially about precipitation forecasts. As a result, climate change mitigation strategies involve anticipating increased climatic variability. In the present case, that means expecting that the present flood frequency is likely to change, perhaps increasing the frequency of inundation but also perhaps decreasing it.

Panel's Draft Response – Effects of Weathering of PCB mixture on Toxicity Issue: Commercial PCB formulations (e.g., Aroclors) contain complex mixtures of many congeners. Once released into the environment, the relative concentrations of congeners in various environmental matrices change based on differences in volatilization, abiotic degradation, adsorption to soil/sediment, biotic metabolism, etc. The BERA emphasized total PCBs when calculating ecological risks because PCBs are regulated on a total and not congener-specific basis in Michigan (page 3-1). Even so, the BERA recognized the potential for weathering of PCB mixtures in the environment (page 4-7) and the relatively greater persistence of PCBs with five or more chlorine atoms (page 4-3). Furthermore, the section on derivation of TRVs for the BERA includes a discussion of both Aroclor-based and TEQ-based risk assessments (Appendix D pp. 26-31). This section of the BERA cites two previous papers from the MSU laboratory (Ludwig et al. 1996, Giesy and Kannan 1998) concluded that "in general, risk assessments based on the original source of Aroclor are likely to underestimate the risk of bioaccumulated PCBs." So while risk estimates and remediation goals may be expressed in terms of total PCBs on account of regulatory requirements, MSU's congener-specific data is a major **strength** that allows the examination of congener patterns, total TEQs, and the contribution of individual congeners to total TEQs. This congener-specific approach reflects the current state of the science in this field.

MSU's conclusion about the weathering of PCB mixtures and reduction in the relative potency of the congener mixtures in Kalamazoo River floodplains (Blankenship et al. 2005) deserves careful examination, especially since this same laboratory has argued that differential weathering, metabolism, and/or bioaccumulation has led to enrichment of toxicity in higher trophic levels of Great Lakes food webs (Giesy et al. 1994a and b, Ludwig et al. 1996 provide several good summaries and reviews; many other papers from the MSU lab also could be cited). The specific arguments using relative potency (RP) factors to demonstrate weathering and reduction of toxicity (Blankenship et al. 2005) are weak. For instance, when examining the great horned owl (Fig. 4 and associated discussion), Blankenship et al. (2005) concluded that RPs show a reduction in toxicity for trophic transfers between small mammals and shrews to owls. Blankenship et al. (2005) downplay the RP of 1.9 for transfer from robins to owls because "it is unclear how much of the great horned owl diet may consist of robins and similar passerine species" (p. 5960). However, Strause et al. (2008) reports that passerines constituted a significant proportion (22%) of the owls' dietary mass at Trowbridge. So in fact there is evidence for enrichment of potency along this important trophic pathway. Furthermore, the two pathways with low RPs in Blankenship et al. (2005) are shown by Strause et al. (2008) to be negligible or irrelevant because they account for only small fractions of the owls' dietary mass at Trowbridge (0.2% for shrews and 6% for other small mammals). RPs for these pathways are meaningless if owls consume few of these food items.

The assertion that BMFs cannot be applied directly to TEQs (Blankenship et al. 2005: 5959) is also highly questionable. The citation of only a single study to support this argument does not reflect the breadth and depth of published studies on TEQs in food webs. Furthermore, BMF considerations for total PCBs and TEQs are extremely similar from chemical and mathematical perspectives. Mathematically, total PCBs are calculated as the sum of individual congener concentrations with a relative weighting of 1 for each congener. TEQs are calculated as a sum using TEFs as weighting factors. Any changes in relative concentrations of individual congeners (i.e., preferential weathering, metabolism, or accumulation of particular congeners) will influence concentrations (and BMFs) of both total PCBs and TEQs.

Panel's Draft Response -- The high concentration of TEQs (both absolute and relative to total PCBs) in house wren eggs, nestlings, and adults is a significant finding and has potential implications for the avian exposure and productivity studies. These elevated concentrations indicate significant exposure to and/or preferential storage of toxic congeners in this species (Table 2.1). Clearly bioaccumulation/retention patterns of toxic congeners vary significantly amongst only the three avian species that were assessed. When considering the diversity of avian species in the Kalamazoo River floodplain, other birds with preferential accumulation and/or retention of planar PCBs may also vary significantly. (Note: the high TEQ concentrations in house wrens are not sufficiently discussed in the MSU papers.)

Table 2.1. Relationship between avian TEQs and total PCBs in birds at the former Trowbridge impoundment

Species/Tissue	Mean avian TEQs (ng/kg)	Mean total PCBs (mg/kg)	Ratio of TEQs/PCBs
House wren eggs	423	8.2	51.6
House wren nestlings	89	1.4	63.6
House wren adults	107	3.2	33.4
Bluebird eggs	57	7.4	7.7
Bluebird nestlings	6.7	1.7	3.9
Robin adults	3.9	0.92	4.2
Great horned owl plasma	0.69	0.49	1.4
Great horned owl eggs	13	7.2	1.8

Data from Blankenship et al. 2005

2.2.2 Question 2: What are the relative strengths, limitations, and uncertainties associated with the productivity assessments conducted by MSU on passerines and great horned owls (Neigh et al. 2006b, 2007; Strause et al. 2007a, 2008).

Panel's Draft Response: Studies of productivity of the bluebirds, wrens, and great horned owls provide useful, qualitative evidence of reproductive performance of on-site species. The **strength** of these studies is that they are directly measuring one of the assessment endpoints ("do PCBs affect reproduction of birds?"). Field measurements are generally preferable to laboratory-based studies as they include much greater realism, including the fact that contaminant-induced changes are not always additive to other stressors that can reduce productivity (weather, predators, etc.). Of course, this is a **limitation** in field studies as well, as it requires a very large sample size to be able to apportion causality to observed effect and to statistically show differences among local populations. The **limitations** of the study are: the small samples sizes (which in some cases are insufficient to draw defensible conclusions); issues with pseudo replication and other aspects of study design; reliance on aquatic organisms for a portion of the diet of the bluebirds and owls; lack of accounting for observational artifacts (such as time of nest initiation or failure) with the great horned owl study; the large effect of one bluebird female's nest failure on the overall success rate of the local population (again reflecting small sample size); the confounding effects of habitat differences between the KRSS sites and the reference site (Fort Custer); and the situation in which the bluebird boxes have been on-site for years at

Fort Custer but were newly erected at Trowbridge (box use is known to be significantly affected by familiarity of the birds with the placement of the boxes).

Supplemental Issues to Consider in Question 2

2a. Strengths and limitations of directly measuring productivity in the field compared to extrapolating from controlled laboratory studies.

Panel's Draft Response: The MSU studies did not characterize population effects because productivity measures alone, without information on survival rates, cannot predict population consequences. In this regard, the assertion that one "bad year" and two "good years" is evidence that contaminants are not affecting productivity is not supportable. There is evidence from other studies that interactions of other environmental stressors (climate and/or parasites) with contaminants can result in reduced productivity, while the same contaminants in otherwise "good years" will not (Nagy, Schumaker, Fairbrother et al., unpublished data on western bluebirds). Furthermore, apportioning cause of nest failure is difficult, but methods now are available to provide a more quantitative estimate (e.g., Etterson et al., 2007).

Panel's Draft Response – Great Horned Owl: The great horned owl is an appropriate receptor species for assessing the risks of PCBs to top predators consuming a mixture of foods from both terrestrial and aquatic food chains. (Note that other top predators exist in the Kalamazoo River floodplain, including red-tailed hawks, other raptors, snakes, and raccoons [the latter two predated bluebirds and house wrens in their nest boxes]). The following analysis does not specifically address extrapolations to other predators. The BERA concluded that the great horned owl was at significant risk to effects of PCBs accumulated through the terrestrial food web. The published MSU papers outline a rationale for monitoring great horned owls in field studies related to ecological risk assessments (Strause et al. 2007a, b, 2008, Zwiernik et al. 2007). However, the amount of data and strength of conclusions in the MSU studies were limited both by characteristics inherent to the biology of this species and by particular aspects related to how the studies were conducted or reported. These **limitations** reduce the ability of the owl dataset to support MSU's associated relative to the evaluation of current and future risk to top predators and to future risk management activities.

Principles of ecosystem energetics dictate that the amount of energy per unit area available to top predators is relatively low, and hence they are constrained to have large home ranges and low population densities. This limits the number of owls that can be supported by the formerly impounded/floodplain areas along the Kalamazoo River, leading to small sample sizes in the MSU studies. Furthermore, not all variables were (or could be) measured at all nests/areas, further reducing sample sizes for some variables, in particular reproductive productivity and call-count indices (see Table 3 of Strause et al. 2007a and Table 1 of Zwiernik et al. 2007 for sampling summaries). Low sample sizes and lack of replication reduced the statistical power for analysis of many variables and eliminated the possibility of inferential statistics for productivity (because n=1 at the Fort Custer reference site).

Further statistical questions are raised because of uncertainties in the temporal—spatial distribution of sampling and the independence of the sampled nests. None of the published papers shows or describes the spatial distribution of sampled nests within a given year or between years. Figure 1.2 in Giesy and Zwiernik (2008) provides a map of owl nest platforms and egg collections at Trowbridge for the all years of the study, but the nests that were occupied and sampled are not designated. Five owl eggs from Trowbridge were collected for residue analysis, but Figure 1.2 shows only 3 egg collection locations. Presumably several eggs were collected from the same nest or territory during different years, but the specifics are not clear. Beyond eggs, presumably some of the same pairs/territories were monitored in multiple years, and hence data would not be independent. Again, these relationships cannot be determined from the papers or presentation.

The assessment of owl reproductive productivity (fledglings per nest) was particularly limited by small sample sizes and lack of replication of sites, thus the published studies over-interpret the strength of this data set. Comparisons of Great Horned Owl populations between Trowbridge and Fort Custer sites cannot be done in any meaningful way due to small sample sizes. Only one active nest was monitored for productivity at the reference site and six at Trowbridge. Several methods have been put forward for calculating nesting and fledging success to account for problems associated with timing of observations, the most common of which is the Mayfield (1975) method. Owl nest productivity might be recalculated using this alternative method to see if additional information about egg production, and hatching or fledging success can be ascertained. However, the data set might be too small, in terms of both sample sizes and the types of collected data, for this reanalysis to be done. Clearly Fort Custer was an insufficient reference site for productivity studies. However, MSU incorrectly concludes (Strause et al. 2007a, Zwiernik et al. 2007) that the "mean" (n=1 for reference) productivity of 1 young/active nest was not "significantly" different between Trowbridge and the reference site.

Furthermore, both of these MSU studies (Strause et al. 2007a, Zwiernik et al. 2007) conclude that Trowbridge productivity was consistent with the productivity found by Holt (1996) in a multi-decade study of 906 great horned owl nets surrounding Cincinnati, Ohio (Holt 1996). This comparison to only *one* other published study on great horned owl productivity is a very limited ecological analysis, and additional studies should be considered. GHO productivity can vary greatly depending on ecological factors, especially food supply. When major prey items such as hares are particularly abundant, productivity can be as high as 2.5-2.6 fledglings per successful nest (Houston 1987 and Houston and Francis 1995, as cited in Holt 1996). Hence, there is little support for the conclusion that a productivity of 1 fledgling per active nest is in fact "normal" for this southwestern Michigan or the upper Midwest.

2b. Extrapolation of results from Site-specific productivity studies to other species such as the robin, which was the receptor species considered in the Final Baseline ERA (CDM, 2003).

Panel's Draft Response: Site-specific studies of species productivity can be extrapolated to other species but only with added uncertainty. Species differences in diets and home range locations would suggest that exposures would differ and, therefore, effects may as well. Compensatory / depensatory factors may have some similarities (e.g., weather – exceptional cold just as young are hatching) or not (e.g., predation; ability to re-nest; size of clutch). On the other hand, species that are in similar feeding guilds (e.g., primarily insectivorous) and of the same type of life history strategy may be sufficiently similar in exposure and interactions with their environment to allow extrapolation of effects from one to the other. However, none of the species studied by MSU (eastern bluebird, tree swallow, house wren, or Great Horned Owl) has diets that are entirely similar to that of the robin. Furthermore, the tree swallow may be particularly insensitive to PCB effects (see below; section 3c), and so would not be a good model for other passerines. Extrapolation of bluebird productivity to robins would be confounded by the fact that bluebirds eat a significant amount of aquatic invertebrates. House wren diets are entirely terrestrial invertebrates but generally do not include earthworms or fruit, both of which are consumed by robins during certain seasons. House wrens have a clutch size of 3-7, while robin clutches are smaller (3-4) and male house wrens tend to be more promiscuous than male robins which may increase productivity per unit area (http://www.birds.cornell.edu). All of these attributes could be accounted for in a qualitative uncertainty discussion if bluebird or house wren productivity measures were to be extrapolated to robins. Given the difficulties with sample size and study design, the added uncertainty reduces the usefulness of such extrapolations.

Panel's Draft Response – Passerines Productivity Methods: Many previous field studies have shown that reproductive and developmental endpoints are useful for assessing risks of PCBs, dioxins, and furans to a variety of avian species, including colonial water birds (gulls, terns, cormorants, and herons), raptors (especially eagles), and passerines (especially tree swallows) Such field studies generally complement the wealth of laboratory studies on the reproductive effects of these chemicals in birds. Hence, the reproductive productivity studies of Kalamazoo River passerines were well-founded and provided significant real-world biological data that were, for the most part, absent from the BERA. However, MSU's productivity studies were limited by several factors, including (perhaps) ecological factors beyond the control of the investigators (e.g., low sample sizes for bluebirds at Trowbridge) but also by issues related to the study design, data analysis, and clarity of data reporting. The use of hypothesis testing is not appropriate with the collected data (see Appendix A). The generally low sample sizes limit the applicability of the conclusions that can be drawn. In particular, the risk of a false negative conclusion (type II error) is high with small sample sizes. Despite the small sample sizes, the data contains a number of indications that reproductive success was lower at Trowbridge compared to the Fort Custer reference area. However, inferences about causation by PCBs are unsupported and would require replication of study areas and (or) additional exposure-response models. Additional analyses might include and analysis of ecological covariates (e.g., habitat characteristics, weather conditions) by building statistical models. Several important spatial-temporal issues related to the study design and data analysis are not reported adequately in the SOPs and publications. (Note: As with the Great Horned Owl studies, the passerine SOPs (260, 262, and 264) seem to reflect initial plans and were not updated to reflect the methods that were actually used. There appears to be no SOPs for bluebirds or house wrens.

Figure 1.2 in Giesy and Zwiernik (2008) provides a map of nest boxes and egg collections at Trowbridge for the all years of the study, but the nests that were occupied and sampled are not designated. None of the published papers or SOPs shows or describes the spatial distribution of sampled nests within a given year or between years. These spatial-temporal relationships are of interest for several reasons. The Stage 1 Assessment for the Kalamazoo River NRDA concludes that different areas of the former impoundments present different levels of risk to terrestrial passerines (see Fig. 7.19 in Stratus 2005). The MSU studies potentially provide the opportunity to examine whether passerine reproduction was affected in the more contaminated sections of the Trowbridge impoundment, but this cannot be done without more information on the location of active nests. More detailed information on the nests also would be useful for censoring and re-categorizing the data. During the May tour of the Trowbridge impoundment by the Panel, MSU scientists described instances in which bluebirds nested in flooded habitats, which would generally be considered atypical for this species and perhaps inappropriate for evaluating risks of PCBs in soils. However, these nests (or any other nests) cannot be identified and removed from the analysis based on the descriptions in the MSU publications. With respect to time, a more complete reporting of initial nests and re-nests would be beneficial. Given the generally equal sample sizes reported for early and late nests (Table 3 of Neigh et al. 2007), these categories seem to reflect an even split of nests rather than a more detailed classification of first and second nest attempts (e.g., based on documented first and second nest attempts in individual boxes).

Another methodological issue was the removal of eggs for chemical residue analysis. This practice introduces significant inaccuracy into the estimates of reproductive success. Consider a nest with 3 viable eggs and 1 nonviable egg. If no eggs are removed and the nest is monitored through hatch, the true hatching success of 75% and brood size of 3 would be known. However, random removal of 1 egg for chemical analysis would result in inaccurate estimates of hatching success. Removal of a viable egg (which would happen with 75% probability) would yield a hatching success of 67% and calculated brood size of 2.75. Removal of the nonviable egg (25% probability) would yield a hatching success of 100% and predicted brood size of 4. Similar considerations exist for numbers of fledglings. MSU's calculations of predicted brood size and predicted number of fledglings cannot fully remove this inaccuracy and variability, nor do they account for biological effects of artificially reduced clutch size (e.g., increased survival of young in manipulated nests because more food is available to each chick). The effects of this inaccuracy were disproportionally greater at Trowbridge. Eggs were removed from 40% of bluebird nests and 39% of house wren nets at Trowbridge, compared to 25% and 20% at Fort Custer (based on sample sizes in Tables 1 and 3 of Neigh et al. 2007). The inconsistent removal rates may have impacted the comparison of hatching success among sites.

Interpretation of variables related to hatching and fledging is often complicated by the definition of these terms, which influences the particular individuals and nests that are included in the calculation of a given variables. Calculation of some of these variables in the MSU studies is not explained clearly. Furthermore, samples sizes often do not agree or add up between Tables 2 and 3 or within Table 3 of Neigh et al. 2007. However, based on the definitions of some of MSU's variables and the trend for declining sample sizes with the advancement of the reproductive cycle (from laying to fledging), it seems that nests that failed early were often ignored for later calculations. For example, fledging success included only "successful" nests in which at least one young fledged, ignoring the biologically relevant nests that failed to produce any young. The consistent decreases in sample sizes from clutch size to hatching success are unexplained. Hatching success should include nests in which no eggs hatched. The predicted number of fledglings is based on an undefined but small (in fact, the smallest) subset of nests. This variable should be calculated for all nests that were initiated. MSU's subdivision of the reproductive cycle into several segments is useful for determining what might be happening during different periods, but it also potentially obscures the greater question of overall reproductive success (i.e., the number of fledglings based on all initiated nests). Incremental effects during various subsections of the breeding cycle might be insignificant in and of themselves but might add up to a larger, more significant effect over the entire reproductive cycle.

To calculate an overall measure of reproductive success using MSU's data, clutch sizes (# of eggs laid/initiated nest) were multiplied by the productivity (number of fledglings/egg laid) to give a fledging rate of number of fledglings per nest initiated (see Table 2.4 below). Fledging rates were 47% lower for bluebirds and 18% lower for house wrens at Trowbridge compared to Fort Custer. Note that removing the Trowbridge bluebird female that experienced repeated failures has little influence on this conclusion. Hence, overall reproductive success appears to have been much lower at Trowbridge, especially for bluebirds.

MSU's presentation of nest success (Table 1 and associated text in Neigh et al. 2007) is also parsed into smaller subsets (years and causes), perhaps missing larger picture differences. Simply summing the data across years (see Table 2.3 below) suggests much lower rates of nest success in bluebirds at Trowbridge and marginally higher rates of nest abandonment in both species at Trowbridge. These variables should be subjected to statistical analysis.

Table 2.2. Coefficients of variation based on Table 3 of Neigh et al. 2007.

	Fort Custer			Trowbridge		
	mean	**sd**	**cv**	**mean**	**sd**	**cv**
<u>Eastern Bluebird</u>						
hatching success	0.79	0.31	39	0.59	0.42	71
fledging success	0.96	0.21	22	0.83	0.35	42
productivity	0.76	0.34	45	0.47	0.44	94
clutch size	4.2	1	24	3.6	1.5	42
predicted brood size	3.8	1.1	29	3.3	1.4	42
predicted number of fledglings						
<u>House wren</u>						
hatching success	0.81	0.25	31	0.64	0.41	64
fledging success	0.92	0.34	37	1	0	0
productivity	0.74	0.3	41	0.64	0.41	64
clutch size	5.7	1.1	19	5.4	1.4	26
predicted brood size	5	1.6	32	4.6	2	43
predicted number of fledglings	4.8	1.6	33	4.6	2	43

Table 2.3 Fledging rates based on all active nests in which clutch size was measured (data from Table 3 of Neigh et al. 2007).

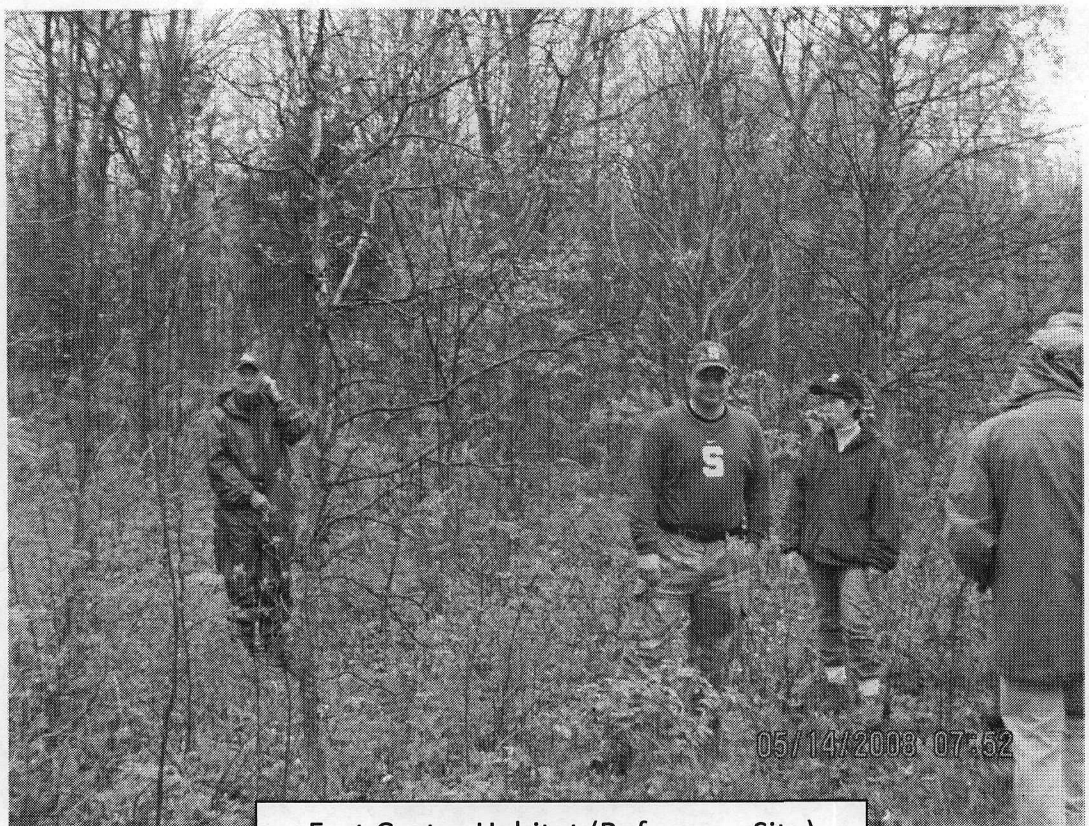
	Fort Custer	Trowbridge	Trowbridge % below Fort Custer	Trowbridge without female that failed repeatedly	Trowbridge % below Fort Custer without female that failed repeatedly
Eastern Bluebird					
Clutch Size (eggs/nest)	4.20	3.60	-14.3	3.60	-14.3
Productivity (# young/egg laid)	0.76	0.47	-38.2	0.51	-32.6
Fledging Rate (# young/nest)	3.19	1.69	-47.0	1.84	-42.2
House Wren					
Clutch Size (eggs/nest)	5.70	5.40	-5.3		
Productivity (# young/egg laid)	0.74	0.64	-13.5		
Fledging Rate (# young/nest)	4.22	3.46	-18.1		

Table 2.4. Nest fate for all years combined (data from Table 2 of Neigh et al. 2007)

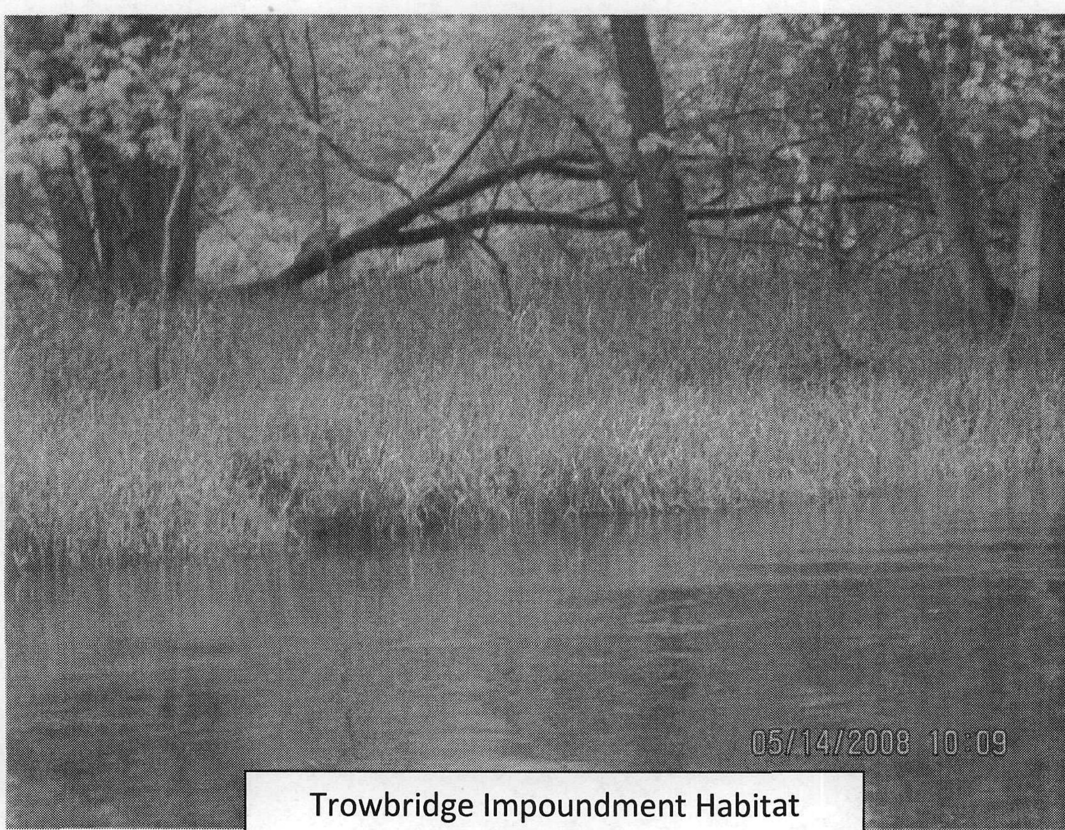
	Fort Custer total	% of Total	Trowbridge total	% of Total
Eastern Bluebird				
Successful	43	75.4	8	44.4
Abandoned	4	7.0	3	16.7
Total	57		18	
House Wren				
Successful	55	77.5	25	73.5
Abandoned	0	0.0	5	14.7
Total	71		34	

2c. Evaluation of potential causal factors (e.g. PCB concentrations, habitat differences, etc) associated with any difference in measures of productivity in passerines relative to the reference site.

Panel's Draft Response: Comparison of avian productivity between KRSS and the reference site (Fort Custer) is difficult because of small samples sizes issues with pseudo replication (e.g., non-representative data) and other aspects of study design, lack of accounting for observational artifacts (such as time of nest initiation or failure) with the great horned owl study, the large effect of one bluebird female's nest failure on the overall success rate of the local population, and the confounding effects of habitat differences among the KRSS sites and the reference area (Fort Custer). Habitat differences include more open areas at KRSS and more riparian woods at the reference site. Complicating the interpretation is that the bluebird boxes have been on-site for years at Fort Custer but were newly erected at Trowbridge; box use (and subsequent contribution to population productivity) is known to be significantly affected by familiarity of the birds with the placement of the boxes(i.e., may be lower at Trowbridge due to unfamiliarity with the boxes). Given these difficulties, it is not possible to make statistical inferences based on hypothesis testing about productivity on KRSS and compared to Ft. Custer.



Fort Custer Habitat (Reference Site)



Trowbridge Impoundment Habitat

2.2.3 Question3: What are the relative strengths, limitations, and uncertainties associated with the hazard quotient calculations performed by MSU to evaluate potential risk to passerines, great horned owls, and shrews (Neigh et al. 2007; Strause et al. 2007a, 2008)

Panel's Draft Response: Hazard quotient calculations consist of two components: a dose estimate and a TRV. MSU's HQs replace many of the modeled doses used in the BERA with site-specific estimated derived from field studies. The use of new site-specific data is an important **strength** of the MSU HQs. The TRV estimates used by MSU, in contrast, were based on the same suite of previously-published studies used in the BERA. Any differences between the TRVs by MSU and the TRVs used in the BERA reflect a combination of differing scientific judgments and differing degrees of conservatism. The Panel can comment on the scientific aspects of MSU's TRV selection process, but defers to USEPA with regard to the appropriate degree of conservatism.

As shown in Tables 2.5 and 2.6 (below), the BERA and MSU approaches differ substantially with respect to both dose estimation and TRV selection. The BERA approach relies heavily on measured concentrations in abiotic media, extrapolated to doses using site-specific and literature-derived transfer factors. The MSU approach relies almost exclusively on PCB concentrations measured in food items. MSU's site-specific approach should provide more realistic estimates of PCB exposures for the purpose of risk assessment; however, transfer factors such as those used in the BERA would still be necessary to calculate remediation goals for soil and sediment.

For passerines, differences between BERA and MSU HQs are driven by the TRVs. There is not much difference between the diet-based MSU and BERA HQs. The TRVs used in the BERA are somewhat lower than those used by MSU, but interchanging the BERA and MSU TRVs changes the HQs by only a factor of 2-3, probably smaller than the uncertainty associated with the original BERA and MSU HQ values. The MSU egg-based TRVs are much higher than either the BERA or MSU diet-based TRVs. Both the NOAEL and LOAEL egg-based TRVs are derived from field studies of nest productivity (tree swallow and robin) rather than controlled laboratory studies. For Great Horned Owl, both assessments use the same TRVs. The differences are due to the large differences in the dose estimate used in the BERA as compared to the MSU studies. The principal contributor to the dose estimate calculated in the BERA (Table C-1) is the estimated concentration of PCBs in robins, which was calculated from earthworm data using a literature-derived biomagnification factor. The MSU dose estimate was based on measured concentrations of PCBs in adult robins.

Table 2.5 Comparison of assumptions and dose calculation methods – BERA vs. MSU

Assumption or method	BERA	MSU
Abiotic media concentrations	U95 confidence bound on arithmetic mean (or max value, if sample size insufficient)	Unknown (used only for incidental soil ingestion)
PCB concentrations in food items	Maximum measured tissue concentrations for earthworms; site-specific soil to plant transfer factors for plant tissues	Measured tissue concentrations in all food items (average?)
Dose estimation	Measured/estimated concentrations in food items + literature-derived diet and standard metabolic parameters from EPA exposure assessment handbook.	Measured concentrations in food items + site specific diet + standard metabolic parameters from EPA exposure assessment handbook.
TRV	Passerines: NOAEL and LOAEL based on chicken (0.4 mg/kg-d to 0.5 mg/kg-d)	Passerines: NOAEL and LOAEL based on pheasant (0.6 mg/kg-d to 1.8 mg/kg-d). Alternative NOAEL and LOAEL extrapolated from egg-based TRVs using biomagnification factors (1.9 mg/kg-d to 14.7 mg.kg-d)
	Great horned owl: NOAEL and LOAEL based on screech owl (0.41 mg/kg-d – 1.2 mg/kg-d)	Great horned owl: NOAEL and LOAEL based on screech owl (0.41 mg/kg-d – 1.2 mg/kg-d)

Table 2.6 Hazard Quotient Calculations Based on MSU or BERA Doses and TRVs

		TRV		
Species	**Dose**	*BERA*	*MSU*	*MSU (egg-based)*
Robin	*BERA*	**1.8-2.3**	0.5-1.5	0.06-0.47
	MSU	N/A	N/A	N/A
Eastern Bluebird	*BERA*	N/A	N/A	N/A
	MSU	**1.0-1.3**	0.3-0.85	0.03-0.3
House wren	*BERA*	N/A	N/A	N/A
	MSU	0.26-0.33	0.07-0.2	0.009-0.07
Great horned owl	*BERA*	**1.8-2.3**	**1.8-2.3**	N/A
	MSU	0.05-0.14	0.05-0.14	N/A

The HQ estimates in Table 2.6 show that, depending on the choice of assumptions and TRV values, HQ values for all three species can be either greater than or less than 1.0. Detailed discussions of the supplemental issues related to HQ development and interpretation are provided below.

Supplemental Issues to Consider

> ***3a. Choice of toxicity reference value (TRV), including relevance to receptor species and quality of study (e.g., duration, inclusion of sensitive life stages, exposure range, endpoints measured).***
>
> ***3b. Uncertainty resulting from extrapolating from laboratory study to field.***
>
> ***3c. Uncertainties in extrapolating from one species to another***

Panel's Draft Response – Toxicity Reference Values: This charge question addresses the appropriateness of MSU's toxicity reference values (TRVs) as well as the use of MSU's data to extrapolate to other species. The Panel is concerned about MSU's approach to develop TRVs for the PCBs and species of concern. The choice of TRVs is a very significant issue that drives risk assessment conclusions. In particular, MSU's TRVs used for calculating the hazard quotients (HQs) (i.e., evaluating risks relative to field-measured or modeled exposures) are quite high and significantly influence their conclusions of no or minimal risk. Application of more protective (but realistic and accepted) TRVs may well lead to different conclusions. Given the critical importance of the TRV selection to the overall risk assessment and the resulting risk management decisions, the detailed rationale and basis used by MSU to derive the TRVs needed to be articulated; the 27 July 2008 memo to the Panel did provide more information on the process used to derive TRVs, but concern remain to the Panel.

Clearly, there should be a technical documentation that describes all of the toxicity literature that was reviewed and the decision rules used to select studies for inclusion in the toxicity database. If a TRV is to be based on a single study, then the specific justification used to select the particular study to represent the effects level is needed, *including the rationale for not selecting alternative studies*. Similarly, if a TRV is derived from a statistical analysis based on several toxicity studies, then the basis for the derivation of the TRV must be explained and justified. The recent memo to the Panel indicated that a single study was selected for establishing each TRV value, but that leaves in question the validity of the TRV that was selected and, consequently, the basis for the risk analysis. It should the emphasized that use of an alternative TRV could lead to the opposite conclusions from those reached by the MSU scientists, as illustrated in Table 2-6.

It also should be noted that the derivation of TRVs for PCBs in birds and mammals is not being done here for the first time. It would seem appropriate for there to be a thorough literature review of TRVs that have been used in other ecological risk assessments involving PCBs in riverine and riparian habitats to provide an indication of the range of avian and mammalian TRVs used in the published literature and other risk assessments. In general, toxicity values may vary greatly across PCB congeners, species, toxicity endpoints, test methodologies, etc., so selecting the particular study or studies to use for defining a TRV needs to be carefully considered, documented, and justified. Indeed, basing a TRV on a single toxicity study seems unwise because of these sources of variability and because the species of concern in the risk assessment are not species used in typical laboratory-based studies, i.e., there must be a reliance on cross-species extrapolations. Calabrese and Baldwin (1993) is a useful reference for TRV derivation methods.

The USEPA issued the final report on the EcoSSL (ecological soil-screening level) methodology last year, and although there is no Eco-SSL document for PCBs, the approach used in developing mammalian TRVs for polycyclic aromatic hydrocarbons (PAHs) would seem to be instructive here (see USEPA 1999, 2005, 2007) In that approach, USEPA examined a large database of published toxicity studies (involving several thousand studies) and applied objective criteria for selecting those particular studies that warranted inclusion in the EcoSSL database for PAHs (about 40 studies met the USEPA criteria). It would seem logical to follow similar criteria in selecting the particular toxicity studies to base TRVs upon. Once the studies were selected, EcoSSL provided a conservative methodology for deriving mammalian TRVs, including, among other things: looking only at population-relevant endpoints (mortality, growth, and reproduction) but not other endpoints (e.g., histological results); including all relevant data in a single assessment to characterize the 95% confidence limits around the geometric means; using the lower CL as the TRV value. Part of the rationale for using this multiple-study, statistical approach with the 95% lower CL as the TRV is to address species differences in toxicity responses. See USEPA (2005) for a detailed flowchart of the steps in derivation of wildlife TRVs. Again, there would seem to be opportunity here to follow a similar approach, basing the TRV derivation as much as possible on USEPA-approved methods. If another process is followed, then there needs to be an explanation of the details of the process that was used and the rationale for following that approach, providing clear justification of the selected TRVs to the exclusion of other potential selections.

The information concerning TRVs that was provided in the publications about how TRVs were selected is too cryptic. For example, in the description by Neigh et al. (2006) on how studies were qualified, there was no mention about congener specificity. Since we know that PCB congeners differ significantly in their toxicity (and mode of action), it is extremely important that previous studies used to set threshold values be done with similar congeners or with toxicity-adjusted equivalents. For example, the ring-necked pheasant study by Dahlgren et al. (1972) used Aroclor 1254, and the initial source of PCBs at the Kalamazoo was Aroclor 1242 (which has since been weathered; see Blankeship et al. 2005). Yet no mention is made about which congeners are contained in these two Aroclors and which ones are likely to have more dioxin-like toxicity. Neigh et al. (2006) also used a study by Nosek et al. (1992) who did an IP injection with TCDD. Neigh et al. (2006) correctly identified the shortcomings of using this study for derivation of a TRV, in particular that the exposure route was via an IP injection. Many will argue that this is an unnatural exposure route and is not suitable for comparison with oral exposures. It would be instructive to hear more justification by the authors for why this study is applicable (is it a "worst case" exposure, thus potentially providing a conservative risk estimate?).

The Panel is also concerned about the use of tree swallow data for determining avian TRVs; this is questionable because tree swallows can accumulate high concentrations of PCBs yet show minimal or no heath/reproductive effects. There may be some taxonomic justification for extrapolating within passerines from tree swallows to bluebirds and house wrens. However, previous avian studies by the MSU lab have used much lower TRVs when considering risks to colonial water birds and bald eagles in the Great Lakes, Giesy et al. (1994 a.b). Considering the diversity of avian species in the floodplain habitats, some of these species are likely to be more sensitive to the effects of PCBs and therefore would be more appropriate for a conservative risk assessment. Note that using the avian TRVs for Aroclors 1242 and 1248 from the BERA assessment (Appendix D) in association with MSU's exposure data would clearly produce HQs above one both for passerine species and owls.

The Panel suggests presenting all of the toxicity studies endpoints on the same graph (similar to that used by the EcoSSLs or as was done by Jim Chapman for the BERA). If these studies are to be used for decision-making, there needs to be a thorough vetting and discussion to reach agreement by all parties on which TRV to use (or, perhaps, several TRVs to bound the probable effects range).

The Panel noted an inconsistency in the MSU TRV development approach in how the lowest-observed adverse effects level (LOAEL) values relate to the no-observed adverse effects level (NOAEL) values. NOAELs cannot always be determined from empirical data since the particular study might not have had a sufficiently low dose to have a result with no significant effects compared to the controls. In that case, NOAELs are occasionally derived from the LOAELs. That was the approach used by MSU, but a multiplicative factor of 3x was used for total PCBs, while a factor of 10x was used for TEQs. Either a rationale for the discrepancy needs to be provided, or else the same factor should be applied.

Neigh et al. (2006) estimated a NOAEL from a reported LOAEL in the pheasant feeding study by dividing the LOAEL by 3, based on an assumption that the reported LOAEL was "near the threshold for effects." They also estimated a NOAEL from a LOAEL for the pheasant IP injection study (using TCDD; Nosek et al. 1992) by dividing by 10, as there were ""pronounced effects" occurring at the LOAEL dose level. This uncertainty should be accounted for by the risk managers when assessing the degree of confidence in the risk outcome. Overall, Neigh et al. (2006) attempted to be highly conservative in their choice of dietary TRVs. Nevertheless, prior to accepting their suggested dietary thresholds, the Panel recommends a thorough and detailed discussion with all parties on the strengths and limitation of the studies in the literature.

Strausse et al. (2007) developed a TRV for the Great Horned Owl based on a study of dietary exposure of a screech owl. The **strengths** of this approach are the taxonomic similarity of the two species and the dietary route of exposure used in the controlled study. The **limitations** are that allometric dose scaling apparently was used (as per Sample et al. 1996), which has been shown to be inappropriate for chronic exposures (Sample and Arenal 1999; Luttik et al. 2005). However, the discussions of allometric adjustments to derive TRVs provided in Strause et al. (2008) for the GHO and in the MSU response to the Panel memo of 25 July 2008 for shrews do not provide details of the scaling factor used, and the language is somewhat ambiguous as to what adjustment was made. If MSU actually just normalized the dose values to a per-weight basis (which is *not* an allometric adjustment), then this is the correct approach to derive chronic TRVs. If, however, MSU actually did use an allometric scaling factor designed to address differences in metabolic rates across species, then this is not correct (and it would not be necessary, anyway, as the actual energetic needs of the GHO are directly known [Duke et al. 1973]). The resulting effect of the allometric approach on the TRV value cannot be determined without knowing what scaling factor was used.

Another **limitation** is that specific PCB congeners were not identified (i.e., to support an argument that the PCBs used in the screech owl study are of greater or similar potency to those at the KRSS). Again, a LOAEL was estimated from the NOAEL using an uncertainty of 3 which is always a subjective approach. The TEQ approach for threshold derivation used the TCDD study by Nosek et al. (1992) and followed the same approach as outline by Neigh et al. (2006).

In summary, the Panel strongly feels that since the selection of the TRV values is critical to determining the results of the HQ-based risk assessments. There must be adequate documentation and justification of the data and the process used to derive the TRVs. The use of a range of TRVs (and, consequently, HQs) in the risk assessment, each fully explained with respect to source and uncertainties, would enhance the utility of the risk analysis and support of the risk management process.

2.2.4 Question 4: What are the relative strengths, limitations, and associated uncertainties that should be considered when evaluating the results of these studies as potential lines of evidence in an Area-specific risk assessment?

Panel's Draft Response: The exposure data collected by MSU should be very useful for quantifying exposures of key ecological receptors addressed in area-specific risk assessments. However, for reasons discussed below, MSU's conclusions concerning risks to passerine birds and great horned owls may not be applicable.

Supplemental Issues to Consider:

4a. Study designs including (but not limited to) sample size, replication, temporal duration, and aggregation of data.

Panel's Draft Response – Conceptual Model: The problem formulation stage of a risk assessment considers a number of important topics, including the specific purposes and scope of the assessment, the choice of receptor species, the identification of critical exposure pathways, the time frame under consideration, and the philosophy of protectiveness (i.e., the adoption of more or less conservative methods for choosing toxicity reference values, making extrapolations between species and locations, etc.). In the risk assessment paradigm, the resulting conceptual model greatly influences the design and execution of subsequent studies as well as the analysis and interpretation of data.

The conceptual models of the BERA and MSU studies differ in significant ways, both positively and negatively influencing the ability of the MSU studies to address questions posed by the BERA and validate and/or revise conclusions of the BERA. While the panel has not been charged with reviewing the BERA, some comments regarding the BERA's conceptual model are helpful for comparison to MSU's conceptual model and assessment.

Overall the intent of the BERA appears to be broadly protective for a wide range of species during both the present and future time frames. Consistent with the conceptual model presented in the BERA, receptor species were selected for study and modeling based a number of criteria, including sensitivity and potential exposure to PCBs. A large number of studies were used to derive TRVs. Some TRVs were intentionally protective to account for the potential presence of sensitive species and extrapolation beyond the modeled receptor species.

By comparison to the BERA, the overall purposes, conceptual model, analyses, and conclusions of the MSU studies were more narrowly focused with respect to species and time and reflect a less protective approach. Further, MSU did not clearly articulate its conceptual model of the potentially important exposure pathways in the areas of concern, nor did they identify how the studies that were conducted contribute to the overall understanding of the ecological risks. The absence of a comprehensive conceptual model of the ecological risks by the MSU team is a major **limitation** that leaves the results of their studies insufficient to challenge some of the conclusions drawn from the BERA.

One major purpose of the MSU studies was to compare conclusions resulting from multiple lines of evidence or different risk assessment approaches (e.g., top-down versus bottom-up approaches). (This objective was an organizing principle for 6 of the 8 published papers and was cited by Dr. Giesy at the May 13 charge meeting as one of two major purposes of their studies. The other major objective was to compare assessments based on total PCBs versus TEQs based on PCB congeners, which was the emphasis of one paper. The 8th paper emphasized other aspects of environmental chemistry in owls.). The MSU studies made this comparison of methods considering only present conditions and only the few species selected for study. Thus, MSU's conclusions do not necessarily apply to all species living in the study area nor to changing future conditions. While the large amount of field data is a **strength** of the MSU studies and the objective of comparing risk assessment methodologies is admirable, the more narrowly defined purpose, lack of a conceptual model, and statistical concerns in the resulting analyses limit the applicability of MSU's conclusions.

The receptor species and exposure pathways chosen for the conceptual model of a risk assessment are critical, and in this case vary significantly between the BERA and MSU studies. The terrestrial receptor species considered in the BERA were muskrats (a linkage between the aquatic and terrestrial food webs), earthworms, deer mice, robins, red fox, wood thrushes, yellow warblers, Great Horned Owls, and red-tailed hawks. A number of terrestrial species were chosen by MSU for ecological studies (earthworms, other terrestrial invertebrates, herbivorous/omnivorous small mammals, shrews, bluebirds, house wrens, and great horned owls). However, the actual studies conducted by MSU emphasized those species that could be sampled easily in the field, primarily through trapping or artificial nesting structures, rather than their potential importance as identified in a comprehensive conceptual model. The choice of these species was sensible given logistical constraints related to obtaining sufficient sample sizes (i.e., abundant species) and accessibility (i.e., easily sampled species), and on account of MSU's main objective to compare lines of evidence derived from ecological sampling and food chain modeling. The inclusion of shrews and terrestrial invertebrates (other than earthworms) in MSU's studies was a significant addition to the BERA. Shrews are likely to experience elevated PCB exposure through their ingestion of earthworms, terrestrial invertebrates, and incidental soil. However, as discussed previously, MSU recently provided the Panel with data from shrew studies, but the limitations in those data, and a the few associated earthworm data, make it unclear at this point how much those studies improve the understanding of the ecological risks. Nevertheless, the Panel does recommend that the shrew data be considered as a component in the multiple-lines-of-evidence approach, both as a pathway for exposure to higher trophic levels and as a representative of the receptor class of small mammals, and conclusions based on the results should be taken as far as the data and uncertainties allow in informing risk management decisions.

Although MSU's studies included to some degree all of the species listed in the BERA for food web/exposure modeling, hazard quotients were calculated only for bluebirds, house wrens, and great horned owls. These species do not necessarily represent the most highly exposed or the most sensitive species present in the riparian corridor. For example, one of the MSU papers acknowledges that the two passerine species "were not chosen as surrogate species or sensitive sentinels for other species but rather were studied to determine the potential for exposure of these species and to determine ecologically relevant reproductive parameters" (Neigh et al. 2007). By failing to characterize additional receptor species, the MSU studies did not focus on all of the important pathways of exposure that a comprehensive conceptual model would have identified. This is a major **limitation** to the conclusions that may be drawn from the MSU studies concerning overall ecological risks.

Likewise, MSU used relatively high avian TRVs, based largely but not exclusively on studies of less-sensitive passerines. While this approach may be appropriate for comparing to multiple lines of evidence for the specific species that were studied, it does not account for potential risk to potentially more sensitive species present in the floodplain ecosystems. Hence, MSU's risk conclusions apply only to the particular species that were studied and are not necessarily broadly protective of entire the avian community.

The soil—earthworm exposure pathway was a particularly important factor in driving a conclusion of potential risk for robins in the BERA. However, the MSU team collected few field data for robins. This was perhaps understandable if, in fact, robins fed rarely in the Trowbridge study area. At the May meeting and in Neigh et al. (2006a), MSU reported a dearth of earthworms in the Trowbridge study area. (Note: this observation is at odds with a) the earthworm sample sizes reported Blankenship et al. 2005 and b) the apparently common occurrence of earthworms reported by CDM for the BERA per DEQ memo.) Earthworms were not found in food bolus samples from bluebirds and house wrens at the Trowbridge site (Neigh et al. 2006a), so ecological studies and hazard quotients related to these two species, while important, provide no information about risks to earthworm-eating birds. Further consideration of the soil—earthworm exposure pathway is essential for evaluating risks to earthworm-eating wildlife such as robins, woodcock, snipe, and shrews.

The delineation of the time frame under consideration is also important to the issue of an adequate conceptual model —the charge asks for an evaluation of MSU's data and analyses as appropriate lines of evidence for "future" risk management decisions. At the May 13 Charge Meeting Dr. Giesy specifically stated that none of the MSU studies were designed to look at future scenarios, such as changes in habitat use with succession. The emphases of MSU's published studies are consistent with that statement, comparing multiple lines of evidence only for the species studied under present conditions. A time scale of many decades is more consistent with the ecological time scales considered by ecosystem managers. This is particularly to be expected if current plans by the State are implemented to completely remove the remnants of the Trowbridge Dam and other water control structures, resulting in altered hydrology in critical areas of concern. This would directly reduce the frequency and magnitude of episodic flooding events, which cause 1) the inundation of the formerly impounded areas with PCB-contaminated sediments, and 2) the erosion and exposure of contaminated soils. An important consequence of such a decreased frequency of inundation may be a significant increase in the density of earthworms in those areas. Thus, earthworm abundance and bioaccumulation of PCBs may change as soils develop over time in the formerly impounded areas, enhancing the magnitude of what was found in the BERA to be the critical pathway for risks. Failure to consider plausible changes in future conditions is a **major limitation** of the MSU studies.

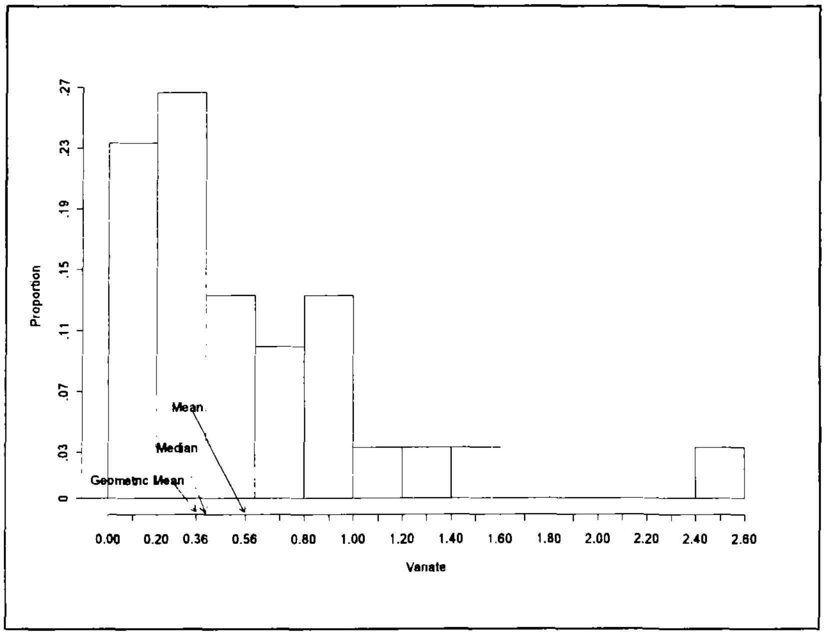
4b. Data interpretation, including the choice and application of statistical methods.

Panel's Draft Response -- Inconsistent and Unclear Design and Analysis Methods: The MSU Sampling and Analysis Plan (SAP) and Standard Operating Procedures (SOP) in Attachment #4 are in conflict with several methods reported in the published papers. One would assume that the methods reported in the published papers are correct; however differences should be explained and justified. For example, the SAP, page 68, calls for stratified random sampling with a random sample of size 2 from each of 3 strata for sediment, soil and biota collection, or 6 locations total. The methods in the published papers indicate that 4 study sites were subjectively selected in the Trowbridge impoundment and 2 were selected in the Fort Custer reference site. Subjective selection of sites was reinforced in oral remarks made by Dr. Zwiernik during the Panel's tour of the Trowbridge impoundment site.

The Panel notes that the null and alternative hypotheses stated in the SOPs should essentially be switched with each other. Fortunately, this error does not appear to have found its way into the published papers or to have influenced the formulas for exploring sample sizes required to detect important effect sizes.

Panel's Draft Response – Inconsistent Statistical Methods: Geometric means were used in HQs for Great Horned Owl studies, whereas apparently, arithmetic means were used for computing HQs for Blue Birds and House Wrens. We assume the upper 95% confidence limit used in computing HQs for the Great Horned Owl is for a confidence interval on the geometric mean. First, methods should be consistent between the papers published by some of the same authors. Second, the geometric mean may not be appropriate for calculation of HQs, because for a sample from positively skewed PCB data, the geometric mean will typically be less than the median and the median is less than the mean, i.e., more than 50% of the sample data will be greater than the geometric mean. The geometric mean provides the smallest value (for positive skewed data) and therefore is least conservative and has the highest chance of a "no risk" decision (e.g., assuming we are talking about an exposure distribution or a distribution of metrics where large numbers are 'bad'). See the attached Figure 2.1, for the location of the geometric mean, median, and mean in an example sample of size 30 from a log-normal distribution. Note that for the log normal distribution, the "most likely" value is estimated by the median. Therefore, we suggest that this value generally be used for all analyses. The arithmetic mean will provide an overestimate of the "most likely" value, and could be used with this explicit understanding (e.g., a conservative estimate of concentration).

Figure 2.1 Geometric mean, median, and mean for a sample of size 30 from a positively skewed log-normal distribution



Panel's Draft Response – Pseudo Replication Issues: It is not possible to make design-based statistical inferences to the entire Trowbridge impoundment and the entire Fort Custer reference site as implicitly implied, because at best, there appears to be a subjectively selected sample of size one from each of four strata in the Trowbridge impoundment and a subjectively selected sample of size one from each of two strata in Fort Custer. Statistical inferences to Trowbridge impoundment and Fort Custer are based on pseudo replication (e.g. resulting in non-representative data) within the relatively small sampling grids and clusters of nest boxes. Statistical results are at best limited to the combined areas of the sampling grids and clusters of next boxes; however, the papers did not appear to check that 'replications' within sampling grids or nest boxes within clusters were far enough apart in space or time to ensure that they were statistically 'independent.' Potential spatial and serial correlation within the sampling grids and within clusters of nest boxes would tend to decrease the effective number of replications and hence increase the standard errors of estimates, increase the width of confidence intervals and increase the p-values of statistical tests, thereby decreasing confidence in results and conclusions by some unknown amount.

Panel's Draft Response – MSU Findings of No Effect: All of the studies have an over dependence on use of statistical tests of null hypotheses. Acceptance of a null hypothesis of 'no difference' between sample sites or times is inappropriately used to conclude that the parameters from the two sites or times are in fact 'equal' or that data sets can be pooled. It is widely known that acceptance of a null hypothesis does not necessarily imply that there are no important biological difference between the parameters under consideration. The summary presented to the

Panel by Drs. Giesy and Zwiernik (both oral and written) at the meeting at Brook Lodge made excessive use of inappropriate acceptance of null hypotheses in attempting to support their conclusions. Furthermore, "no effect" is dependent upon sample size and variability in the data; thus, one might argue that a more robust design with larger sample size (particularly for GHO and blue birds) could return different results.

4c. Approach for addressing natural variability

Panel's Draft Response: There are two aspects of natural variability that are relevant here: within-site variability and between-site variability. Within-site variability refers to variability of exposures and effects within a specific impoundment, e.g., Trowbridge. Between-site variability refers to variability of exposures and effects between impoundments. The MSU studies did not directly evaluate either type of variability; however, data collected by MSU could easily be used to analyze within-site variability. Information on PCB concentrations in food items, eggs, and avian tissues are summarized in the MSU papers as means and standard deviations. All further analyses (e.g., dose modeling and HQ calculations) are then based only on the means. It would be relatively straightforward to use these data to quantify the variability of exposures within either of the two sites studied. For example, the means and standard deviations of PCB concentrations in prey items could, through Monte Carlo analysis of the dose model, be used to compute the distribution of doses within the bluebird and house wren populations inhabiting the Trowbridge and Fort Custer sites. From these distributions, the fraction of birds potentially exposed above any given level (e.g., all of the selected TRV dose) could be computed (General Electric Company, 2005). Similarly, estimates of the means and variances of PCB concentrations and eggs could be used to compute distributions associated with the distributions of resulting egg concentrations. A complete probabilistic assessment could be done by using a range of TRVs rather than a single number, and using Monte Carlo techniques to compare the range of exposure values with the entire range of TRVs. This would essentially negate the need for deciding which single TRV to use, and instead will compute the HQ probabilities associated with each combination of exposure dose and effect level.

Since only two sites were studied by MSU, relatively little information is available about between-site variability. As noted elsewhere in this review, information on between-site differences in diet compositions could be obtained by comparing diets of birds nesting at the Trowbridge and Ft. Custer sites. This information could provide a bound on between-site variability of diets at other sites, because the Ft. Custer site appears to be ecologically quite different from any of the contaminated impoundments.

4d. Identification and characterization of uncertainties.

Panel's Draft Response: Neither the BERA nor MSU studies conducted a formal uncertainty analysis in their respective approaches to establishing ecological risks. Given the differences in risk characterization among the studies, the Panel strongly recommends that the BERA and MSU data have formal uncertainty analyses performed. MSU, for example, could generate exposure and effects distributions using probabilistic techniques (rather than the simple hazard quotients). Similarly, simple sensitivity analyses of hazard ratios formed with differing estimates of exposure and effect could be implemented and graphically presented. (See Appendix A for discussion of approaches to conducting uncertainty analyses). Other sources of uncertainty have been discussed by the Panel elsewhere in its responses, with the overall conclusion by the Panel that there has been insufficient identification and characterization of uncertainties, and insufficient attention to the implications of the uncertainties to the weighing of results and conclusions that could be drawn form the studies.

4e. Adequacy of the data to support inferences on population-level effects.

Panel's Draft Response: The MSU productivity studies did not measure all the parameters necessary to develop a population model and make predictions about long-term changes in per area density of the various bird species (either as individuals or as breeding pairs). This apparently was not the intent of the studies, which were designed to specifically address questions about PCB-related changes in productivity. It frequently is inferred that a statistically significant reduction in productivity will lead to population declines, but this is truly an untested hypothesis for each species. Compensatory or density-dependent changes in juvenile or adult survival may off-set reduced productivity, resulting in similar densities over time. Of course, the age structure of the population could change to an older aged population if productivity, but not survival, were affected. The **strengths** of the MSU studies are the site-specific productivity metrics that were developed (for some of the species). These potentially could be used in a simple matrix population model, using stage-specific survival data from the literature (preferably from studies in the same geographic area). The **limitations** of the MSU data are the small sample sizes and study design difficulties. Those data (such as for the bluebirds) that are sufficiently robust to provide a reliable estimate of fledging success per nest would be very useful in simple population models; this will allow a sensitivity analysis to be conducted to confirm which life stages(s) are most affected and how this will impact the population over the long term.

Panel's Draft Response: The assertion that one "bad year" and two "good years" is evidence that contaminants are not affecting productivity is not supportable. There is evidence from other studies that interactions of other environmental stressors (climate and/or parasites) with contaminants can result in reduced productivity, while the same contaminants in otherwise "good years" will not (Nagy, L, N. Schumaker, and A. Fairbrother, unpublished data on western bluebirds in the Willamette Valley, OR). Furthermore, apportioning causes of nest failure is difficult, although methods now are available to provide a more quantitative estimate (e.g., Etterson et al., 2007). A good example of a true demographic (population) study of a passerine is a study of the western bluebird by Keyser et al. (2004).

Call counts provided indices of population density for Great Horned Owls, but these indices may have been influenced by the use of artificial nest platforms. This habitat manipulation significantly reduces the strength of conclusions (e.g., no effects of PCBs) that can be drawn from these population indices.

The study of metapopulation dynamics moves beyond population density to examination of recruitment rates and landscape-level spatial patterns (e.g., source/sink population dynamics). In ecotoxicology, the key question related to metapopulation dynamics is whether young raised in that area survive and reproduce, either in that area or elsewhere. A previous study by the MSU lab showed significantly low recruitment rates of young Caspian terns raised in ecosystems contaminated with PCBs and other organochlorine contaminants (Mora et al. 1993). Likewise, the MSU lab showed that bald eagles nesting on contaminated Great Lakes shorelines had reduced productivity of young, which contributes to a lower recruitment rate from those areas (Best et al. 1994). In both situations, the numbers of breeding adults in the contaminated areas were supported by the immigration of birds raised in less contaminated areas. The Kalamazoo River studies did not assess recruitment rates and migration patterns, and hence potential effects of PCBs on metapopulation dynamics remain unknown.

Panel's Draft Response – Great Horned Owl Population Estimates: The nest productivity data address organism-level effects and do not support inferences concerning population-level effects of PCB exposures. The great horned owl call count surveys conducted by MSU potentially address population-level effects; however, the Panel believes that MSU's conclusions concerning those effects are not supported by the data.

Call count surveys of great horned owls were conducted for two purposes: 1) locating owl nests for studies of tissue residues, reproductive productivity, and dietary composition, and 2) as indirect measures of relative population density. Two of the MSU papers use the call counts as lines of evidence of a higher owl population density at Trowbridge compared to the Ft. Custer reference, supporting a conclusion of lack of population-level effects at Trowbridge (Strause et al. 2007a, Zwiernik et al. 2007). While this conclusion based on call count indices is consistent with the higher number of occupied nests at Trowbridge, inconsistencies and incomplete explanations in MSU's descriptions of the call count methods raise significant questions about specific (i.e., numerical) comparisons with previously published and potential future studies. The SOP (272) and published papers appear to be inconsistent with respect to cited references and methodological description. The SOP cites Frank (1997) as the source of the methods. Strause et al. (2007a) cites Brenner and Zarkowski (1985) and Zwiernik et al. (2007). Zwiernik et al. (2007) cites both Frank 1997 and Brenner and Karwoski 1985 and possibly Rhoner and Doyle 1992). These citation questions might seem irrelevant but for the different descriptions of the methods in the SOP and papers. The SOP describes evening surveys with call locations at 0.5 km intervals and pre-broadcast, broadcast, and post-broadcast periods at each location. Strause et al. (2007a) describes an "active" method (active not defined) with morning and evening surveys at call locations at 0.5 km intervals. Zwiernik et al. (2007) describes both an active survey method with hoot broadcasts and a passive or silent survey during sensitive life stage events (these events not defined but Rhoner and Doyle 1992 is cited). Given the variability in the above descriptions, it is difficult to determine how data for various observation methods and periods were used to calculate the response rates in the categories of "total," "foraging," and "paired" used in both MSU papers. (Note that MDEQ's concerns over potential biases in the call-count methods are difficult to evaluate because the description of MSU's methods are so unclear.)

Overall, these inconsistencies raise the questions of exactly what was done and whether the protocols were compatible with well-accepted methods and previously published data for great horned owls. A specific example of such a (potential) comparison is the statement in Zwiernik et al. (2007) that "measures of site-use (abundance) indicated the target area populations at Trowbridge were near the carrying capacity for undisturbed GHO habitats (Houston et al. 1998)." This statement is unclear whether the "site-use (abundance)" comparison is being made based on 1) call count indices of relative abundance (responses per survey) generated using identical field protocols in both studies, or 2) estimates of breeding population density (number of breeding pairs per unit area) using call counts and many other observational methods to identify all breeding pairs in an area. In either case, identical or compatible methods would have to be used in both studies to allow specific numerical comparisons to be made.

Beyond these issues of call count protocol, the usefulness of the relative abundance indices is severely limited by the insufficiency of the Ft. Custer as a reference site. Call counts using the exact same methods should be applied to other (and replicated) references sites in southwestern Michigan or the upper Midwest to determine the magnitude of and variability in these indices in healthy populations.

2.2.5 Question 5: What are the relative strengths, limitations, and associated uncertainties that should be considered when extrapolating from the results of MSU studies conducted in the former Trowbridge Impoundment to the other formerly impounded areas of the Kalamazoo River?

Panel's Draft Response: Given the limited amount of habitat- and spatial-related information provided to the Panel, there is considerable uncertainty in extrapolating results from the Trowbridge impoundment to other formerly impounded areas. The MSU papers and reports do not adequately describe the relationship between habitat and exposure/effects data. A spatial re-analysis of MSU's data might be insightful, but in some cases the samples sizes for particular types of samples are quite small or data were composited across sampling locations, severely limiting their use in habitat-specific analyses and extrapolations. While the types of habitat and plant communities found in the former impoundments appear to be generally similar, they do appear to differ in their relative distribution, and perhaps in other important characteristics including patch size and connectedness, which may affect the conclusions that could be drawn from the studies. A more definitive description of future land management goals by the Trustees would also help clarify these extrapolation questions.

Supplemental Issues to Consider:

5a. Numeric and spatial distributions of PCBs in floodplains of former impoundments

Panel's Draft Response: The numeric and spatial distribution of PCBs in soil and biota is poorly described in MSU's papers and reports. Fig. 1-2 in MSU's May workshop report (Giesy and Zwiernik 2008) shows soil/biota sampling and nest box/platform locations, but no subsequent information is provided linking sampling locations with numeric PCB data. (Presumably such information exists in MSU's database, but the clarity of this information has not been examined by the Panel.) Given the low sample sizes for some sample types, conclusions regarding spatial (and temporal) variability are likely to be significantly limited. Note that Chapter 7 of the Stage 1 NRDA assessment (Stratus 2005) does a more complete job of characterizing risk on a spatial basis and may provide an example of what could be done with MSU's data (when sufficient).

5b. Habitat characteristics in floodplains of formerly impounded areas

Panel's Draft Response: Factors to be considered when making these extrapolations include variations in habitat type, food web structure, soil/sediment PCBs, and likely utilization by various passerine species. At a minimum, impoundment-site-specific conceptual models will be needed to identify the key uncertainties relevant to each impoundment. To account for future changes in diet composition, these conceptual models should also include changes in habitat characteristics related to ecological succession.

The obvious ecological differences between Fort Custer and Trowbridge could be used to characterize between-site differences in passerine diets. The MSU publications combine diet data for each species over all sites. Although this approach is useful for comparing the diets of different species over a range of habitats, it obscures within-species differences in diets that may occur due to differences in habitat quality or prey availability at different sites. Within-species comparisons of diet compositions of birds nesting at Trowbridge to diet compositions of birds nesting at Fort Custer would permit at least a qualitative evaluation of the influence of site characteristics on passerine diets. Supporting reference: Neigh et al. (2006), which contains diet composition data for tree swallow, house wren, and eastern bluebird, for Trowbridge and Fort Custer sites combined.

Note: The Panel's ability to respond to Charge 5b has been limited by the availability of site specific habitat-related information for the former impoundments. The MSU papers and reports contain little information on the spatial distribution of sampling, including the relationships between samples and particular habitat types. Habitat information from the Trustees has also been minimal. Some of the most detailed information on habitat is found in Section 3 and Figures A-1 through A-5 in ARCADIS, 2008 report provided to the Panel at the May 13 Charge Meeting. The Panel was told that this information had not been vetted by the Trustees. Examination of the habitat maps shows generally similar habitat types present in the former impoundments. Comparisons of the degree of habitat fragmentation are difficult given the different scales used for some habitat maps.

5c. Likely utilization of floodplains in formerly impounded area by the receptor species evaluated in MSU studies

Panel's Draft Response: Since the mix of habitats at Trowbridge seems to be generally similar to the mix at the other impoundments, one would expect to find similar receptor species present. The particular receptor species chosen for study by MSU are relatively common species for this region of Michigan, and hence would be expected to be present if appropriate habitat is available. Great horned owls might be one exception--the smaller impoundments would likely be big enough for only a few (or even a partial) owl territory. Populations of cavity nesting passerines such as bluebirds and house wrens would be limited by the abundance of natural cavities (e.g., in dead trees) in the absence of nest boxes. Relatively simple, qualitative wildlife survey's (e.g., visual bird observations or call counts, limited small mammal trapping) could be used to clarify uncertainties in the animal community composition at the other impoundments.

2.2.6 Question 6: Please comment on the applicability of the information presented in the MSU studies for informing risk management decisions.

Panel's Draft Response: The applicability of the information presented in the MSU studies for informing risk management decisions depends upon the following four considerations:

- Data quality: conformance to USEPA standards for sample collection/handling, analytical chemistry, database management, etc.
- Study design: species/site selection, selection of metrics, sample size, as function of study objectives.
- Relative value of empirical studies performed by MSU (soil/biota concentrations; nest productivity; analysis of PCBs in nestling diets) vs. literature-based analyses (TRVs)
- Interpretation of results: Conclusions supported by data?

With regard to data quality, it appears to the Panel that MSU followed USEPA's recommended procedures. With regard to study design, as noted elsewhere in this review, MSU's approach was significantly narrower than the approach taken in the BERA; moreover, one of the key receptors (robin) evaluated in the BERA was not addressed by MSU. In addition, the nest productivity studies were compromised by small sample size, pseudo replication, and lack of comparability between the Trowbridge site and the Fort Custer reference site. The other empirical studies performed by MSU, specifically the measurements of PCB concentrations in soil and biota, the direct measurements of PCB doses to nestlings, and the quantification of species-specific dietary preferences, did not suffer from these flaws. These empirical studies provide new data that can be used to support refined area-specific risk assessments and other studies performed to support the risk management process at this site. MSU's approach to TRV selection does not appear to be superior to the approach used in the BERA and provides no new information for risk management. Because conclusions concerning risks presented in MSU's published papers are heavily dependent on values chosen for TRVs and the justification for selecting specific TRVs is inadequately described, the Panel believes that MSU's risk conclusions are not supportable. However, the risk assessment approach used in the BERA could be modified to accommodate MSU's site-specific exposure data, thereby significantly enhancing the quality of risk information available to risk managers.

In this regard, the Panel notes that neither the BERA nor MSU implemented any sort of formal uncertainty analysis in their respective approaches to establishing risk. Again, given the differences in risk characterization among the studies, the Panel strongly believe that formal uncertainty analyses should be conducted to support any future use of the MSU data and for any other data used in risk management. For example, exposure and effects distributions could be generated using probabilistic techniques (rather than the simple hazard quotients). In addition, simple sensitivity analyses of hazard ratios formed with differing estimates of exposure and effect could be implemented and graphically presented. Uncertainty bounds on the resulting clean-up values could be generated. A comprehensive listing of mathematical approaches for conducting these analyses is not presented here, but can be found in Warren-Hicks and Moore (1998) and Warren-Hicks (1999).

3.0 Panel's General Comments

The studies conducted by Dr. Giesy and his team provide additional valuable information to inform the ecological risk assessment and risk management decisions on the Kalamazoo River. The **strength** of their work lies in the site-specific data collected that can be used to verify the dietary exposure models used in the BERA and provide additional lines of evidence. The additional lines of evidence include egg concentrations of PCBs and productivity measures of the study species. The **limitations** of the studies include: inadequate statistical design; insufficiency of many of the data; absence of a comprehensive conceptual model relating exposures and effects on endpoints of concern; the lack of detailed information in the publications (i.e., the lack of a study report that could contain much more detail than allowed in a literature paper); inadequate documentation and justification of the selected TRVs; inadequate identification and quantification of sources of uncertainty; and the over-interpretation of the results provided in Dr. Giesy's summary document.

The best use of the MSU study result would be to:

- Use the site-specific tissue data in the dietary exposure models in the BERA – the **strength** of this approach is to provide site-specific BSAFs and BMFs and measured concentrations in biota, rather basing the food web model entirely on literature based estimates. This will incorporate soil-specific effects (e.g., soil carbon), congener-specific differences in accumulation rates, and species-specific information related to the site (particularly for raptors, where literature-based data are very sparse). The **limitation** is that PCB concentrations in earthworms were inadequately measured, so the robin exposure pathway cannot be verified. Therefore, the estimate in the BERA will need to stand as the best assessment of risk to robins, although it may be modified/strengthened by site-specific adjustments of BSAFs used to estimate earthworm concentrations.

- Use the "bolus" data from the avian nesting study to further verify dietary exposure estimates – the **strength** of this approach is that the food bolus represents precisely what the nestlings are eating. By comparing the concentrations in this bolus to the estimated concentrations from the dietary exposure model, the model can be further refined to accurately reflect the diets and exposures (BSAFs/BMS) of the studied species. This may provide some additional realism for extrapolating to the non-measured species, such as the robin. The **limitations** of this approach is that only the house wren is truly feeding on only terrestrial foods, while the eastern bluebird, the tree swallow, and the great horned owl access some (or most) of their diets from the aquatic food chain. Thus, relating diet to soil contamination alone will be difficult.
- Use the great horned owl dietary assessment (Strausse et al., 2008) as the input to the exposure assessment for raptors at KRSS. A **strength** of this study is the direct measurements of PCBs in some GHO prey items that rarely are analyzed, and a reasonable comparison between KRSS and the reference site. However, it should be pointed out that PCBs were not measured by the MSU studies in rabbits and large squirrels represent 50 to 75% of GHO diet based on mass basis. Further strengths are the presentation of data on both a mass-basis and a concentration basis, plus inclusion of both means and 95% UCLs of the means. However, until agreement is reached on appropriate TRVs, the hazard assessment presented in the paper should not be used.
- Studies of productivity of the bluebirds, wrens, and great horned owls provide useful, qualitative evidence of reproductive performance of on-site species. The **strength** of these studies is that they are directly measuring one of the assessment endpoints ("do PCBs affect reproduction of birds?"). Field measurements are preferable to laboratory-based studies as they include much more realism, including the fact that contaminant-induced changes are not always additive to other stressors that can reduce productivity (weather, predators, etc.). Of course, this is a limitation in field studies as well, as it requires a large sample size to be able to apportion causality to observed effect and to statistically show differences among local populations. The **limitations** of the study are the small samples sizes, issues with pseudo replication and other aspects of study design, reliance on aquatic organisms for a portion of the diet of the bluebirds and owls, lack of accounting for observational artifacts (such as time of nest initiation or failure) with the Great Horned Owl study, the large effect of one bluebird female's nest failure on the overall success rate of the local population, and the confounding effects of habitat differences among the KRSS sites and the reference area (Fort Custer). Further complicating the interpretation is the bluebird boxes have been on-site for years at Fort Custer but were newly erected at Trowbridge; box use is known to be significantly affected by familiarity of the birds with the placement of the boxes. Nevertheless, these studies can be used in a

qualitative manner to relate site productivity with generally expected reproductive success of the species within the region.

- The MSU data can also be used to build models linking measurements of dietary intake to body burden. In addition, the MSU data can be used to evaluate relationships between measurement endpoints over space. Each of these approaches should provide insights currently not found in the BERA.

These studies should not be used to reach risk conclusions on their own. There is too much uncertainty underlying the data interpretation, lack of robustness in the study design, and insufficient documentation (and lack of agreement) of TRV derivation. Some of the papers are repetitive (e.g., Strausse et al., 2008 and Zwiernik et al., 2007 both describe risk to Great Horned Owls using essentially the same data). One study (Strausse et al. 2007) on relationship of PCB concentrations between nestling blood plasma and eggs in great horned owls and bald eagles is interesting and provides good information for future monitoring studies, but is not particularly relevant to the current risk assessment at KRSS. Otherwise, the papers each contribute some data and information that can be used in an assessment of risk if integrated with the data and approaches used in the BERA.

The work by Giesy and his team provide useful data for quantitative exposure estimates and qualitative weight of evidence for estimating effects. They can contribute information to provide a more robust assessment of risk than currently provided in the BERA, but not as stand-alone documentation.

In summary, the work by Giesy and his team provide useful data for quantitative exposure estimates and qualitative weight of evidence for estimating effects. They can contribute information to provide a more comprehensive assessment of risk than currently provided in the BERA. However, the limitations of these studies indicate that that should not be used as stand-alone documentation, and the conclusions presented based on these data are not supportable.

4.0 Panel's Draft Recommendations – Looking Forward

1. The Panel recommends that a cross-comparison between the MSU and BERA studies be made using data from one in the model from the other. For example, it would seem to be a simple exercise to take the PCB concentration data in soils and lower-trophic level samples and run them through the BERA model to see what exposures to higher-trophic-levels would ensue. Similarly, the BERA data could be run through the MSU exposure model. The Panel has already performed a preliminary cross-comparison, as shown in Table 2.6 in this report. It is to be expected that the results will differ between the two studies, because one is primarily literature-based and the other primarily field-studies-based. Nevertheless, such a cross-comparison could illustrate the magnitude of the differences resulting from the two approaches and the causes of the differences. If considered along with an improved understanding of uncertainties associated with each approach, this could then inform the risk managers of the validity and defensibility of each set of analyses, and enhance their ability to

appropriately weigh differences in results. This exercise would seem to be essential before any reasonable understanding of the multiple-lines-of-evidence approach could be reached from the two disparate datasets and result. Moreover, the MSU data, with measurements on the endpoints of interest, should not be limited to analysis using only the USEPA approaches. See Appendix A for other approaches, as well as additional quality assurance issues, that could be applied to the MSU data sets. The MSU data, with measurements on the endpoints of interest, should not be limited to analysis using only the EPA approaches. See Appendix A for other approaches, as well as additional quality assurance issues, that could be applied to the MSU data sets

2. The Panel recommends that, rather than focus on estimating a single risk number, the risk assessments would be strengthened by presentation of a distribution of risk levels tied to the uncertainties in the underlying data and/or model structure (e.g., relative importance of different dietary pathways). Consequently, the Panel recommends that the exposure models and data in the MSU study and the BERA be subject to a formal uncertainty analysis. Included in this should be an extensive sensitivity analysis of the models to explore the plausible range of risks in the system. For example, one set of the variables in an exposure model is the particular diet of an endpoint species. The frequency distribution of dietary sources could readily be varied across a large number of scenarios, allowing calculation of how sensitive the resulting assimilated dose is to the dietary composition. Similarly, use of different specific bioaccumulation factors within the range of plausible values for each could be explored in a set of Monte Carlo simulations. Other model structural and parameter sensitivity analyses would enhance the understanding of the ecological risks in this system and could suggest specific additional research needed to reduce uncertainties.

3. The panel recommends an explanatory model-based approach to data evaluation over the calculation, and re-calculation, of uncertain hazard quotients up an ecological pathway tree. See Appendix A for an explanation of this approach.

4. To address concerns about the approach used in the MSU study to develop toxicity reference values (TRVs) for the PCBs of concern, the Panel recommends that consideration be given to applying an approach that uses the full set of available, high quality toxicity studies rather than a single study to derive a TRV. Such an approach could be modeled after the EcoSSL methodology USEPA developed for PAHs, or after the methodology used by USEPA (2004) to develop the TRVs used in the Housatonic River BERA. Regardless of what method is used, the Panel recommends that the derivation of the TRVs is explicitly described and documented, and the specific TRVs selected fully justified to enhance the confidence that the TRVs that are selected are appropriate and protective. This is critical, since the selection of TRVs directly affects the risk assessment conclusions. A more useful approach would be to use a range of plausible TRV values for each receptor of concern, enhancing the utility of the results to the multiple-weight-of-evidence approach for risk management.

5. The Panel recommends that all parties explicitly consider the time horizon appropriate for the risk assessment and the risk management, and that they give due consideration to the long-

time dynamics of PCB uptake, sediment dynamics and ecological succession, all of which directly affect the results of a risk assessment.

6. Because the distinction between the terrestrial and aquatic habitats of the system seems to be an area of uncertainty and thus is an issue on how to appropriately address the reality of the floodplain dynamics, the Panel recommends that the models used in the BERA and by MSU to calculate dose to ecological endpoints be run in a mixture of terrestrial and aquatic modes.

7. The Panel recommends analyzing the avian reproduction data using the Mayfield method (or similar approach; Mayfield, 1975; Johnson, 1979). This would account for differences in time of observation relative to nest initiation and other similar factors. It is standard practice in most avian productivity studies.

5.0 References

Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin. 2004. *Bayesian Data Analysis*. 2nd Edition. Chapman & Hall, New York, NY. pps: 668.

ARCADIS. 2008. Baseline Ecological Risk Assessment Peer Review Scope of Work for the Allied Paper, Inc. /Portage Creek/Kalamazoo River Superfund Site. March 2008.

Best, D.A., W.W. Bowerman, IV, T.J. Kubiak, S.R. Winterstein, S. Postupalsky. M.C. Shieldcastle. And J.P. Giesy. Jr. 1994. Reproductive Impairment of bald eagles along the Great Lakes shorelines of Michigan and Ohio. Pp. 697-702 in Raptor Conservation Today, Proceedings of the IV World Conference on Birds of Prey and Owls. B.-U. Meyburg and R.D. Chancellor, eds. Robertsbridge, U.K.: Pica Press.

Blankenship, A.L., M.J. Zwiernik, K.K. Coady, D.P. Kay, J.L. Newsted, K. Strause, C. Park, P.W. Bradley, A.M. Neigh, S.D. Millsap, P.D. Jones, and J.P. Giesy. 2005. Differential accumulation of polychlorinated biphenyl congeners in the terrestrial food web of the Kalamazoo River Superfund site, Michigan. Environmental Science and Technology 39:5954-5963.

Brenner, F.J., and D. Karwoski. 1985. Vocalizations as an index of abundance for the great horned owl (*Bubo virginianus*). Proceedings of the National Academies of Science 59:129-130.

Burnham, K.P., and D.R. Anderson. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Second Edition.* Springer, New York.

Calabrese EJ, Baldwin LA. 1993. *Performing Ecological Risk Assessments.* Boca Raton, FL: Lewis Publishers. 257 pp.

CDM, 2003. Final (Revised) Baseline Ecological Risk Assessment – Allied Paper, Inc./Portage Creek/Kalamazoo River Superfund Site. Prepared on behalf of the MDEQ Remediation and Redevelopment Division. April 2003.

Chow, Shein-Chung, and Jen-pei Liu. 2008. *Design and Analysis of Bioavailability and Bioequivalence Studies, Third Edition.* Chapman & Hall/CRC Biostatistics Series.

Duke, G. E., J. G. Ciganek, and O. A. Evanson. 1973. Food consumption and energy, water, and nitrogen budgets in captive Great Horned Owls (*Bubo virginianus*). *Comparative Biochemistry and Physiology* 44A: 283–292.

Efroymson, R. A., G. W. Suter II, B. E. Sample, and D. S. Jones. 1997. Preliminary remediation goals for ecological endpoints. ES/ER/TM-162/R2, Oak Ridge National Laboratory, Oak Ridge, TN.

Etterson, M.A., L.R. Nagy, and T.R. Robinson. 2007. Partitioning risk among different causes of nest failure. The Auk. 124:432-443.

Etterson, M.A., L.R. Nagy, and T.R. Robinson. 2007. Partitioning risk among different causes of nest failure. The Auk 124:432-443.

Frank, R.A. 1997. Effects of dieldrin on great horned owls at the Rocky Mountain Arsenal National Wildlife Refuge. M.S. Thesis. University of Wisconsin, Madison, WI.

General Electric Company (GE). 2005. *Interim Media Protection Goals Proposal for Housatonic River, Rest of River.* Can be downloaded from:http://www.epa.gov/NE/ge/thesite/restofriver-reports

Giesy, J. and M. Zwiernik. 2008. Ecological consequences of PCBs in the exposed sediments of formerly impounded areas of the Kalamazoo River. Presentation to the Kalamazoo River Ecological Risk Studies Peer Review Panel, 13 May 2008.

Giesy, J., and K. Kannan. 1998. Dioxin-like and non-dioxin-like toxic effects of polychlorinated biphenyls (PCBs): Implications for risk assessment. Critical Reviews in Toxicology 28:511-569.

Giesy, J.P., J.P. Ludwig, and D.E. Tillitt. 1994a. Deformities in birds of the Great Lakes region: Assessing causality. Environmental Science and Technology 28:128A-135A.

Giesy, J.P., J.P. Ludwig, and D.E. Tillitt. 1994b. Dioxins, dibenzofurans, PCBs and colonial, fish-eating water birds. Pages 249-307 in A. Schecter, ed. Dioxins and Health. Plenum Press, New York.

Hayward, J.L., J.G. Galusha, and G. Frias. 1993. Analysis of great horned owl pellets with rhinoceros aukelt remains. The Auk 110:133-5.

Holt, J.B., Jr. 1996. A banding study of Cincinnati area great horned owls. J. Raptor Research 30:194-197.

Houston, C.S. 1987. Nearly synchronous cycles of the great horned owl and snowshoe hare in Saskatchewa. Pp. 56-58 in R.W. Nero, R.J. Clark, R.J. Knapton, and R.H. Hamre (eds). Biology and conservation of northern forest owls: symposium proceedings. Gen. Tech. Rep. RM-142, U.S. Dept. Agric., Forest Service, Fort Collins, CO.

Houston, C.S., and C.M. Francis. 1995. Survival of great horned owls in relation to the snowshoe hare cycle. Auk 112:44-59.

Houston, C.S., D. G. Smith, and C. Rhoner. 1998. Great horned owl. Pp. 1-27 in A. Pool and F. Gill (eds). Birds of North America No. 372. The Birds of North America, Inc., Philadelphia, PA

IPCC (Intergovernmental Panel on Climate Change). 2007. *Climate Change 2007.* Four Volumes: a) *Synthesis Report.* b) *Working Group I Report, The Physical Science Basis.* c) *Working Group II Report, Impacts, Adaptation, and Vulnerability.* d) *Working Group III Report, Mitigation of Climate Change.* Intergovernmental Panel on Climate Change, United Nations Environmental Programme and World Meteorological Organization. Available at: http://www.ipcc.ch/index.htm.

Johnson, D.H. 1979. Estimating Nest Success: The Mayfield Method and an Alternative The Auk, Vol. 96, No. 4 (Oct., 1979), pp. 651-661.

Keyser, A.J. J.T. Keyser, D.E.L. Promislow. 2004. Life-history variation and demography in western bluebirds (*Sialia mexicana*) in Oregon. The Auk 121:118-133.

Ludwig, J.P., H. Kurita-Matsuba, H.J. Auman, M.E. Ludwig, C.L. Summer, J.P. Giesy, D.E. Tillitt, and P.D. Jones. 1996. Deformities, PCBs, and TCDD-equivalents in double-crested cormorants (*Phalacrocorax auritus*) and Caspian terns (*Hydroprogne caspia*) of the upper Great Lakes 1986-1991: Testing a cause-effect hypothesis. Journal of Great Lakes Research 22:172-197.

Luttik, R., P. Mineau and W. Roelofs. 2005. A review of interspecies toxicity extrapolation in birds and mammals and a proposal for long-term toxicity data. Ecotoxicology. 14: 817-832.

Mayfield, H.F. 1975. Suggestions for calculating nest success. The Wilson Bulletin Vol. 87, No. 4 pp. 456-467.

Moore, D. R. J., B. E. Sample, G. W. Suter, B. R. Parkhurst, and R. S. Teed. 1999. A probabilistic risk assessment of the effects of methylmercury and PCBs on mink and kingfishers along East Fork Poplar Creek, Oak Ridge, Tennesssee, U.S. A. Environmental Toxicology and Chemistry 18:2941-2953.

Mora, M.A., H.J. Auman, J.P. Ludwig, J.P. Giesy, D.A. Verbrugge, and M.E. Ludwig. 1993. Polychlorinated biphenyls and chlorinated insecticides in plasma of Caspian terns: Relationships with age, productivity, and colony site tenacity in the Great Lakes. Archives of Environmental Contamination and Toxicology 24:320-331.

MSU, 2001

MSU, 2008 Overview of Studies Conducted by Michigan State University presented to the Review Panel at the meeting held at Brook Lodge, May 13 and 14, 2008

Neigh, A.M., M.J. Zwiernik, A.L. Blankenship, P.W. Bradley, D.P. Kay, M.A. MacCarroll, C.S. Park, P.D. Jones, S.D. Millsap, J.W. Newsted, and J.P. Giesy. 2006a. Exposure and multiple lines of evidence assessment of risk for PCBs found in the diets of passerine birds at the Kalamazoo River Superfund site, Michigan. Human and Ecological Risk Assessment 12:924-946.

Neigh, A.M., M.J. Zwiernik, A.L. Blankenship, P.W. Bradley, D.P. Kay, M.A. MacCarroll, C.S. Park, P.D. Jones, S.D. Millsap, J.W. Newsted, and J.P. Giesy. 2006a. Exposure and multiple lines of evidence assessment of risk for PCBs found in the diets of passerine birds at the Kalamazoo River Superfund site, Michigan. Human and Ecological Risk Assessment 12:924-946.

Neigh, A.M., M.J. Zwiernik, C.A. Joldersma, A.L. Blankenship, K.D. Strause, S.D. Millsap, J.L. Newsted, and J.P. Giesy. 2007. Reproductive success of passerines exposed to polychlorinated biphenyls through the terrestrial food web of the Kalamazoo River. Ecotoxicology and Environmental Safety 66:107-118.

Neigh, A.M., M.J. Zwiernik, P.W. Bradley, D.P. Kay, P.D. Jones, R.R. Holem, A.L. Blankenship, K.D. Strause, J.L. Newsted and J.P. Giesy. 2006a. Accumulation of Polychlorinated Biphenyls (PCBs) from Floodplain Soils by Passerine Birds. Environmental Toxicology and Chemistry, Volume 25, pp. 1503-1511.

Rhoner, C., and F.I. Doyle. 1992. Methods of locating great horned owl nests in the boreal forest. Journal of Raptor Research 26:33-35.

Sample, B. E., and C. A. Arenal. 1999. Allometric models for interspecies extrapolation of wildlife toxicity data. *Bulletin of Environmental Contamination and Toxicology* 62: 653-663.

Stratus Consulting, Inc. 2005. Stage I Assessment Report, Volume 1 — Injury Assessment: Kalamazoo River Environment Final Report. Michigan Department of Environmental Quality, Michigan Attorney General, U.S. Fish and Wildlife Service, and National Oceanic and Atmospheric Administration.

Strause, K.D., M.J. Zwiernick, S.H. Im, J.L. Newsted, D.P. Kay, P.W. Bradley, A.L. Blankenship, L.L. Williams, and J.P. Giesy. 2007b. Plasma to egg conversion factor for evaluating polychlorinated biphenyl and DDT exposure in great horned owls and bald eagles. Environmental Toxicology and Chemistry 26:1399-1409.

Strause, K.D., M.J. Zwiernick, S.H. Im, P.W. Bradley, P.P. Moseley, D.P. Kay, C.S. Park, P.D. Jones, A.L. Blankenship, J.L. Newsted, and J.P. Giesy. 2007a. Risk assessment of great horned owls (*Bubo virginianus*) exposed to polychlorinated biphenyls and DDT along the Kalamazoo River, Michigan, USA. Environmental Toxicology and Chemistry 26:1386-1398.

Strause, K.D., M.J. Zwiernik , J.L. Newsted, A.M. Neigh, S.D. Millsap, C.S. Park, P.P. Moseley, D.P. Kay, P.W. Bradley, P.D. Jones, A. L. Blankenship, J.G. Sikarskie, and J.P. Giesy. 2008. Risk assessment methodologies for exposure of great horned owls (*Bubo virginianus*) to PCBs on the Kalamazoo River, Michigan. Integrated Environmental Assessment and Management 4:24-40.

Strause, K.D., M.J. Zwiernik , J.L. Newsted, A.M. Neigh, S.D. Millsap, C.S. Park, P.P. Moseley, D.P. Kay, P.W. Bradley, P.D. Jones, A. L. Blankenship, J.G. Sikarskie, and J.P. Giesy. 2008. Risk assessment methodologies for exposure of great horned owls (*Bubo virginianus*) to PCBs on the Kalamazoo River, Michigan. Integrated Environmental Assessment and Management 4:24-40.

Suter, G.W. II. 2007. Ecological risk assessment. CRC Press, Boca Raton, FL. 643pp.

U.S. Environmental Protection Agency (USEPA) 1993. Wildlife Exposure Factors Handbook. USEPA Office of Research and Development. EPA/600/R-93/187)

U.S. Environmental Protection Agency (USEPA). 2004. *Ecological Risk Assessment for General Electric (GE)/Housatonic River Site, Rest of River.* Can be downloaded from: http://www.epa.gov/NE/ge/thesite/restofriver-reports

US Environmental Protection Agency (USEPA). 1999. *Screening Level Ecological Risk Assessment Protocol. Appendix E: Toxicity Reference Values.* USEPA Office of Solid Waste. Washington, DC.

US Environmental Protection Agency (USEPA). 2005. *Guidance for Developing Ecological Soil Screening Levels.* OSWER Directive 9285.7-55. USEPA Office of Solid Waste and Emergency Response. Washington, DC.

US Environmental Protection Agency (USEPA). 2007. *Ecological Soil Screening Levels for Polycyclic Aromatic Hydrocarbons.* Interim Final OSWER Directive 9285.7-78. USEPA Office of Solid Waste and Emergency Response. Washington, DC.

Warren-Hicks, W. J. 1999. Formal Methods for Risk-Based Decision-Making. HERA 5(2):225-229.

Warren-Hicks, W. J., D. Moore. eds. 1998. *Uncertainty Analysis in Ecological Risk Assessment: Pellston '95.* SETAC Press, Florida.

Zwiernik , M.J., K.D. Strause, D.P. Kay, C.S. Park, A.L. Blankenship, and J.P. Giesy. 2007. Site-specific assessments of environmental risk and natural resource damage based on great horned owls. Human and Ecological Risk Assessment 13:966-985.

6.0 Appendix A - Suggested Statistical Analyses and Quality Assurance Evaluations

Before results between the BERA and MSU studies can be compared and contrasted, a basic assumption concerning the comparability of the data used in each study is required. The MSU study and BERA reach different conclusions. However, the reason for these differences is not readily apparent, but could be associated with such concerns as (1) data collection methods, (2) geographical differences in sampling locations, (3) differences in analytical chemistry methods, (4) differences in mathematical and statistical methods, (4) changes in biota or PCB concentrations over time, (5) selection of toxicity threshold values, etc. We strongly suggest that a rigorous comparison be generated that illustrates the basic comparability of data collected in each of the studies. Below is a brief listing of a few of the many issues that should be addressed and presented to the reader in an understandable format:

- With the exception of Table 1-1 of the MSU Summary Document, a direct easy to read and understand comparison of the data characteristics associated with each study is unavailable to the reader. We strongly encourage such a table(s) (or figure) be developed. At a minimum, the comparable information should include sample location, number of samples, type of data collected, collection dates, etc. associated with each study.

- In keeping with the first bullet, we suggest a similar one-to-one comparison table be created for the data analysis methods employed by each study. At a minimum, the reader should be able to easily juxtapose such information as TRVs used in the dietary and tissue hazard quotients for each endpoint, equations for back calculating cleanup values, average PCB concentrations (or TEQs) used in the numerator of the hazard quotients sorted by location, time, endpoint, trophic transfer factors used to calculate bioaccumulation and clean up values, etc. Again, the reader should be provided easy access to information that will allow further investigation on how or why the two studies reach such different conclusions.

- Neither report addresses the comparability of the most basic data element, concentration of PCBs in various media. If indeed, a soil sample collected by MSU results in a different concentration than a replicate sample collected by EPA, then a comparison of statistical outputs using the two independent data sets is compromised. Therefore, the reader must be convinced that the studies are actually evaluating the same concentration information (including PCB concentrations, TEQs, etc).

- The effect of time is not addressed by either study. It seems that among the two studies, data are collected over many years. The investigators seem to have ignored the role that time effects can play on the comparability of the two data sets and associated findings. A rigorous evaluation of the effect of time on the discrepancy in the study findings should be implemented.

Recommended Statistical Analyses: For the simpler parameters, analyses should be based on point estimation of parameters with measures of precision such as standard errors and confidence intervals. Confidence intervals should be computed for the difference or ratio of parameters and plotted graphically so that potentially important biological differences can be seen. Alternatively, box and whisker plots of data collected from two different sites or times can be displayed in graphical form side by side. Unfortunately, the authors cannot use acceptance of null hypotheses to justify final conclusions without further qualifications.

Consider Table 2-1 in the Overview of Studies Conducted by Michigan State University presented to the Review Panel at the meeting held at Brook Lodge, May 13 and 14, 2008, and Neigh et al. (2007, number 4 in the papers provided to the review team). See bottom of page 110 and page 111 in Neigh et al. (2007). For example, small sample sizes within a year or acceptance of a null hypothesis of 'no difference among years' are not justification for combining reproductive data of eastern bluebirds or house wrens among all years. The decision to pool data from different sources is a subjective decision, not a statistical inference. Confidence intervals could be computed for each year of the reproductive parameters and plotted next to each other on the same figure to provide useful information concerning the differences and degree of variation within each year. Alternatively, box and whisker plots adjacent to each other will provide essentially the same information. Generally, models are fitted to data from various sources in time and space. That is, in addition to the plots over time and sites, multiple regression models could be fitted to explore the relationships between reproductive parameters and predictor (independent) variables such as: year, early versus late nests, sites within Fort Custer and within Trowbridge, Fort Custer versus Trowbridge, etc. Tests of hypotheses and measures of precision associated with the models are subject to question because of the pseudo replication in these studies. Granted that the statistical inferences are limited, useful models may be obtained. ANOVA was conducted using linear models in some cases, however only for the inappropriate purpose of testing null hypotheses. Models for prediction of observed effects on parameters, as functions of covariates measured on the study sites and times, should be developed using variations of the *Akaike's information criterion* (AIC) for selection among competing models (see, e.g., Burnham and Anderson 2002).

In short, tests of null hypotheses should probably not be used in any of the MSU statistical analyses, a realization that is beginning to take root in many disciplines of science. If tests of hypotheses are to be used in evaluating impacts of PCBs or other toxicants then they should be stated in terms of 'tests of bioequivalence' (see for example, Chow and Liu 2008).

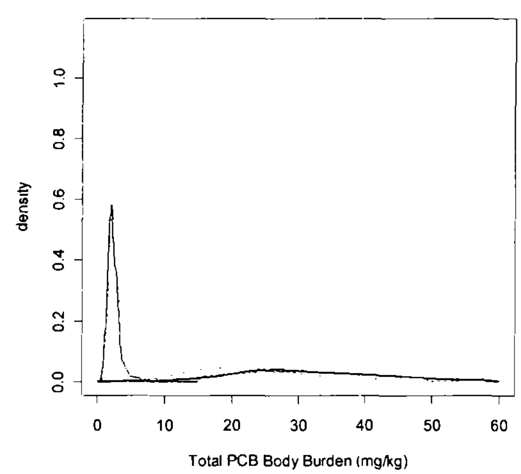
Development of Explanatory Models: The MSU investigators did not take advantage of the relatively rich data set they collected. Boiling the data down to simple hazard quotients is under use of a valuable and rich data set. The great advantage of the MSU data relative to the BERA data is that MSU had data on the response endpoints of interest and site-specific measures of PCB uptake. Granted there are problems with pseudo replication, MSU is afforded the ability to explore relationships among the data variables, examine the distribution of these variables over time and space, and use this information to draw valuable inferences on the PCB exposure potential and the relationship of PCB concentrations to effects.

As an example, the following model was used in the analysis of PCB effects on the Housatonic River (http://www.epa.gov/region1/ge/thesite/restofriver-reports.html#Eco).

$$T_{BodyBurden} = (MT + \sum_{i=0}^{i=14}(FIR_i \ x \ C_{diet} \ x \ 1day) \ / \ BW_{14}$$

$$FIR(kg \, / \, kg \ bw \, / \, day) = \frac{\alpha \ BW^{\beta}}{K}$$

We believe that MSU collected data on PCB body or tissue burden (T), egg concentration (MT), concentration in diet (C), and body weight (BW). These types of models should be used to explore relationships among the endpoints of interest (e.g., body or tissue burden among others) and co-variables (egg concentration, diet, etc.) in an effort to generate additional insights that are achievable beyond hypothesis testing or simple hazard quotients. For example, we believe that the MSU has data for calibrating the above model at several locations within Trowbridge and in the reference area. If so, then information like that generated at the Housatonic River (below), through the model, can be used as an alternative to hypothesis testing to infer the magnitude, uncertainty, and geographic differences in the endpoints of interest.



In the above figure, the red dotted lines represent the distribution of PCB body burden at each of three reference sites, and the solid red line represents integration across all reference sites. Notice that the distributions are tall and thin with similar centers indicating little uncertainty among and within the reference sites. In contrast, the blue lines indicate the impact sites which show a great deal of uncertainty about the center of the distribution, but are easily seen to be different on average than the reference sites. Such probabilistic/graphical analysis of the data provides an insightful way of generating inferences from the data without the use of hypothesis testing.

Co variables like maternal transfer (egg concentrations); can be treated in the same way (see above figure). And, measures of "so what" (i.e., the effects endpoints) can be overlain on such exposure distributions to aid in decision-making. In fact, if distributions of effects (not employed by either EPA or MSU) are available, then an integration of the exposure and effects distributions may prove useful. The integration can be preformed graphically, or numerically.

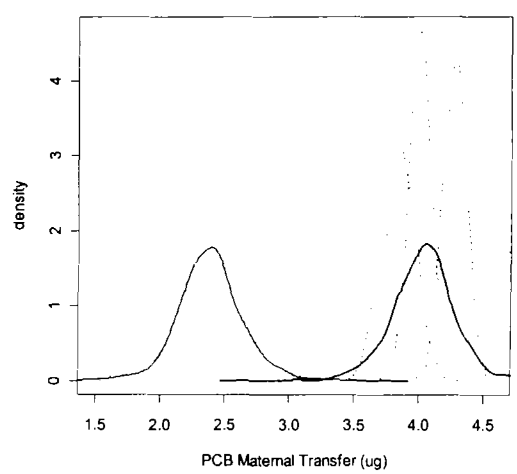
We believe that other mathematical/probabilistic model forms can be devised and implemented for the variety of data types and variables represented by the MSU data sets.

Also note that such models and graphical outputs can be used as an aid toward resolving extrapolation issues. Suppose we wish to extrapolate the results found at one site to another. For example, suppose we have measurements of the model response variable at Trowbridge, but not downstream at Otsego. However, if co variable measurements are available at Otsego, then statistical methods exist that will allow the extrapolation of the Trowbridge information, through the model, to estimate the expected body (or tissue) burdens at Otsego (see Gelman, et al. 2004).

In any case, because the MSU investigators generated substantially different findings than BERA a more extensive analysis of the data should be conducted by MSU. And, given that the MSU data information content is richer than that collected by EPA; we believe that additional data analysis approaches should be used to expand the insights available from the information.

Formal Uncertainty and Sensitivity Analyses for Confirming Risk

Neither BERA nor MSU implemented any sort of formal uncertainty analysis in their respective approaches to establishing risk. Again, given the differences in risk characterization among the studies, we strongly believe that both EPA and MSU should conduct formal uncertainty analyses. MSU, for example, could generate exposure and effects distributions using probabilistic techniques (rather than the simple hazard quotients). And/or, simple sensitivity analyses of hazard ratios formed with differing estimates of exposure and effect could be implemented and graphically presented.



BERA could easily generate uncertainty bounds on the resulting clean-up values.

At the end of the day, investigators associated with the BERA and MSU studies should examine the degree of overlap among the competing data sets and analytical outputs, and decide whether or not the findings are significantly different within the bounds of the available information content represented by the collected metrics.

A comprehensive listing of mathematical approaches for conducting these analyses is not presented here, but can be found in Warren-Hicks and Moore (1998) and Warren-Hicks (1999).

Pooling Data Across Studies: The MSU summary states:

1. *The information was collected with the primary goal of addressing uncertainty in the Baseline ERA (CDM, 1999) ...*

2. *Our intent was to develop multiple, site-specific, independent lines of evidence to supplement those evaluated in the Baseline ERA ...*

Given the above goals, particularly the notion that the MSU data were collected to supplement the BERA evidence lines, we strongly believe that BERA and MSU should attempt to merge the data sets, or at least provide a one-to-one evaluation of the measured metrics in each study. There are possibly a number of advantages to pooling the information across studies, including creating a longer time-series of information, increased information content, increased geographical scale, and an increased ability to draw inferences based on the data information content.

There are a number of methods for pooling data including the following: (1) simple concatenation of data sets using expert judgment to identify those cases where the data cannot be pooled based on scientific reasoning, (2) formal methods for pooling based on underlying probability distributions, and (3) updating approaches when the data are time-dependent (see Gelman 2004). However, there are many measures, such as tissue concentrations and body burdens that BERA generated from literature values and MSU measured in the field. In these cases, the field measurements are preferred and datasets should not be combined.

The analyses described above should only be implemented after an evaluation of the ability to pool data from the studies is completed.

Time-Series Analyses: Data used in the MSU analyses were generally collected from 2001 – 2003. Little or no information is provided in the BERA on the time spans over which data were collected and compiled for the various analyses employed by BERA (see, for example Table C-1 and notice that data characteristics including data collection times are not provided).

The investigators should convince the reviewers and users that data collection time is not a factor underlying the discrepancy in risk characterization results among the studies. Time-series plots, using data from both studies (see above comment), should be generated. Hypothesis testing should not be used as a basis for pooling data over time (see above comment). If specific metrics are shown to have time trends or cycles, the effect of this observations on the risk characterization results must be described, and specific mathematically defensible methods for formally incorporating a time component into the risk analyses must be implemented.

Tiered Risk Assessment: Both MSU and BERA have effectively employed simplistic statistical and data analytic approaches for evaluating the data, typically those employed during the early tiers of a formal risk analysis. Given the discrepancy in risk characterization among the studies, more advanced statistical and risk characterization techniques (like those described above) are warranted. In particular, we encourage MSU and BERA to reduce the dependence of the risk decisions on risk quotients, and implement techniques that make full use of the available information. In particular, uncertainty analyses, time-series evaluations, descriptive graphical analyses, and explanatory models should be used to further evaluate the data and provide insights into the differing risk decisions generated by the BERA and MSU.